

یک الگوریتم جدید برای ساختار مرز طبقات در جوامع چوله

سید مسعود حسینی محمودی^۱، عباس گرامی^۲

چکیده

برای تعیین مرز طبقات در نمونه گیری های طبقه بندی شده روش های متعددی وجود دارد. در این مقاله، ضمن معرفی روش هندسی^۳ برای ساختن طبقات به مقایسه این روش با روش لاوالیه- هایدروگلو^۴ و دالینوس- هاج^۵ که از روش های رایج در تعیین مرز طبقات هستند می پردازیم. با استفاده از شبیه سازی، نشان می دهیم که روش هندسی ضمن ساده بودن دارای خواص بهینه نسبت به دو روش فوق است.

واژه های کلیدی: طبقه بندی، روش هندسی، چولگی، ضریب تغییرات، تخصیص، بهینگی.

۱- مقدمه

برای اولین بار دالینوس روشی را برای تعیین مرز طبقات ارائه کرد، سپس آنرا با همکاری هاج [۲] تکمیل کرد. این روش بر پایه جمع ریشه دوم فراوانی ها استوار بود. و به دلیل سادگی مورد اقبال آمار دانان قرار گرفت و هنوز استفاده از آن رایج است. سپس اکمان [۳] روش دیگری ارائه کرد که از بسط سری تیلور در محاسبات آن استفاده می شد ولی بعلت محاسبات زیاد و فراهم نبودن ابزار های محاسباتی که در حال حاضر موجود می باشند مورد توجه قرار نگرفت. اساس کارهای او ساختن طبقاتی بود که دارای ضریب تغییرات تقریباً مساوی باشند. کوکران [۱] ایده برابری ضریب تغییرات درون طبقات را ارائه کرد. لاوالیه و هایدروگلو [۸] با یک الگوریتم چند مرحله ای نشان دادند که برابری ضریب تغییرات منجر به بهینه ترین طبقه بندی می شود، این الگوریتم

با هدف مینیمم کردن مقدار نمونه بر اساس یک مقدار ثابت دلخواه خطا طراحی شده است. گانینگ^۶ و همکاران [۶،۵،۴] و هرگان^۷ [۷] الگوریتم هندسی را ارائه کردند، که بر برابری ضریب تغییرات در درون طبقات استوار بود. در این مقاله سعی داریم این روش را با ارائه مثال هایی شرح دهیم.

۲- نمونه گیری طبقه بندی

فرض کنیم جامعه مورد بررسی به L طبقه، طبقه تقسیم بندی شده است بطوری که طبقه h ام شامل N_h عضو است. از هر طبقه نمونه ای به اندازه n_h ($1 \leq h \leq L$) انتخاب کرده ایم. قرار می دهیم:

$$n = \sum_{h=1}^L n_h \text{ و } N = \sum_{h=1}^L N_h,$$

در نمونه گیری طبقه بندی شده برآورد میانگین از فرمول زیر حاصل می شود:

$$\bar{y}_{st} = \sum_{h=1}^L \frac{N_h}{N} \bar{y}_h \quad (2-1)$$

که در آن \bar{y}_h میانگین نمونه در طبقه h ام است.

می دانیم که:

^۱ دانشکده ریاضی، آمار و علوم کامپیوتر-پردیس علوم - دانشگاه تهران

^۲ دانشکده ریاضی، آمار و علوم کامپیوتر-پردیس علوم - دانشگاه تهران

^۳ Geometric progression

^۴ Lavellée, and Hidirolou

^۵ Dalenius, T. and Hodges, J.

^۶ Ekman

^۷ Cochran

^۸ Gunning

^۹ Horgan

$$k_h^* = k_{h+1} k_{h-1},$$

این رابطه در محاسبه جملات تصاعد هندسی در ریاضی استفاده می شود. بنابراین خواهیم داشت:

$$k_h = ar^h; \quad h = 0, \dots, L, \quad (3-6)$$

$$a = k_0, \quad r = \left(\frac{k_L}{k_0} \right)^{\frac{1}{L}},$$

که در آن k_0 مینیمم مقدار و k_L ماکزیمم مقدار متغیر کمکی مورد بررسی است.

برای نحوه انجام محاسبه در یک مثال عددی فرض کنید:

$$L = 4, \quad k_0 = 1, \quad k_L = 10000,$$

در این صورت در ۶-۲ داریم:

$$k_h = 1 \times 10^h; \quad h = 1, 2, 3, 4,$$

و دامنه طبقات بصورت زیر خواهد بود:

$$1-10, \quad 10-100, \quad 100-1000, \quad 1000-10000$$

مثال فوق به وضوح سادگی روش موجود را نشان می دهد.

همانطور که بیان شد این روش بر اساس یکنواخت بودن توزیع در درون طبقات استوار می باشد و برای توزیع های دیگر مانند توزیع نرمال زیاد مناسب عمل نمی کند. مگر آنکه تعداد طبقات بسیار زیاد باشد که این خود مشکلات دیگری را ایجاد می کند.

۴- نگاهی به روش های دالینوس- هاج و

لاوالیه- هایدر و گلو

روش دالینوس- هاج [۲]، روشی تقریبی بوده اما دارای نتایج خوبی است. در این روش ابتدا داده ها دسته بندی شده و سپس مقدار ریشه دوم فراوانی ها در هر دسته محاسبه می شود. سپس مقدار تجمعی این ریشه های دوم را مثل جدول تجمعی جمع نمونه و در نهایت مقدار پایانی را به تعداد طبقات مورد نظر تقسیم کرده و بر اساس آن طبقات مشخص می گردد. فرض اساسی در این روش، یکنواخت بودن تقریبی توزیع داده ها در درون طبقات است.

روش لاوالیه- هایدر و گلو [۸] برای طبقه بندی، نیازمند استفاده از توزیع دقیق متغیر کمکی در جامعه تحت بررسی است. اما مشخص کردن توزیع دقیق مشاهدات چندان ساده نیست، بعلاوه این روش دارای یک الگوریتم ۴ مرحله ای همراه با

$$V(\bar{y}_{st}) = \sum_{h=1}^L \left(\frac{N_h}{N} \right)^2 \left(1 - \frac{n_h}{N_h} \right) \frac{S_h^2}{n_h} \quad (2-2)$$

که در آن $S_h^2 = \frac{1}{N_h} \sum_{i=1}^{N_h} (y_{hi} - \bar{y}_h)^2$ واریانس و

تا این واریانس را مینیمم کنیم.

در این مقاله فرض بر آن است که توزیع جامعه از لحاظ متغیر مورد بررسی چوله به راست است و سعی بر آن است تا روشی بسیار ساده را برای تعیین مرز طبقات در راستای هدف فوق ارائه شود.

۳- روش هندسی

فرض کنید بخواهیم جامعه ای را به L طبقه با نقاط پایانی $k_0 < k_1 < \dots < k_L$ برای متغیر Y بر اساس متغیر کمکی X تقسیم کنیم. فرمول ضریب تغییرات متغیر کمکی مذکور برای طبقه h بصورت زیر است:

$$CV_h = \frac{S_h}{\bar{X}_h}; \quad h = 1, \dots, L, \quad (3-1)$$

برای برابری ضریب تغییرات باید تساوی زیر برقرار باشد:

$$\frac{S_1}{\bar{X}_1} = \frac{S_2}{\bar{X}_2} = \dots = \frac{S_L}{\bar{X}_L}. \quad (3-2)$$

در این الگوریتم نیز مانند روش دالینوس فرض بر آن است که توزیع متغیر مورد نظر در طبقات بطور تقریبی یکنواخت است، حال با در نظر گرفتن توزیع یکنواخت در درون طبقات برای متغیر کمکی داریم:

$$\bar{X}_h = \frac{k_h + k_{h-1}}{2}, \quad (3-3)$$

$$S_h = \frac{1}{\sqrt{12}} (k_h - k_{h-1}),$$

با قرار دادن (۳-۳) در (۳-۱) داریم:

$$CV_h = \frac{(k_h - k_{h-1}) / \sqrt{12}}{(k_h + k_{h-1}) / 2}. \quad (3-4)$$

برای برقراری تساوی بین ضریب تغییرات درون طبقات، تساوی زیر را خواهیم داشت:

$$\frac{k_{h+1} - k_h}{k_{h+1} + k_h} = \frac{k_h - k_{h-1}}{k_h + k_{h-1}}, \quad (3-5)$$

و با انجام اعمال ریاضی داریم:

مقیاس ۱۰ و شکل ۱۰ است، جامعه دوم (rst) از توزیع t -چوله با پارامترهای مکانی ۲۰، مقیاس ۱۰، شکل ۱۵ و درجه آزادی ۳ است و جامعه سوم (exp(rsn)) نمایشی داده های تولید شده از توزیع چوله نرمال با پارامترهای مکانی ۵، مقیاس ۱ و شکل ۱ است، هیستوگرام آنها به ترتیب در ادامه آمده است.

برای انجام مقایسه ها ابتدا بر اساس مقدار n مشخص و با استفاده از روش دالینوس- هادج و روش هندسی ساختار طبقات تعیین گردیده است. با توجه به میزان CV بدست آمده از اجرای طبقه بندی با دو روش بالا، این مقدار به عنوان میزان دقت در الگوریتم لاوالیه- هایدروگلو جایگزین گردیده و مقدار نمونه بهینه حاصل شده است. تلاش شده تا در مقدار CV یکسان مقدار حجم نمونه در روش های مورد بررسی، مقایسه شوند. نتایج در جداول ۲-۵، ۳-۵ و ۴-۵ جمع آوری شده است. در این جداول مقدار CV از تقسیم جذر واریانس حاصل از طبقه بندی بر میانگین کل بدست می آید.

تکرار بوده که به مراتب انجام تعیین مرز طبقات را با سرعت کمتری روبرو می کند و گاهی این الگوریتم همگرا نشده و نتیجه ای از طبقه بندی حاصل نمی گردد. در این روش به ازای مقدار خاصی از دقت، مقدار مینیمم n محاسبه می گردد. میزان دقت بر اساس مقدار CV بدست آمده از اعمال روش مورد نظر بر روی جامعه بدست می آید.

حال در بخش بعد با استفاده از شبیه سازی روش های فوق الذکر را با هم مقایسه می کنیم. در اینجا لازم است ابتدا بصورت مختصر سه توزیع چوله به راست را که در این شبیه سازی مورد استفاده قرار گرفته اند معرفی شوند.

۵- شبیه سازی و مقایسه روش های مختلف

برای انجام مقایسات میان سه روش گفته شده در مقدمه یعنی روش دالینوس- هاج، لاوالیه- هایدروگلو و روش هندسی از شبیه سازی با نرم افزار R استفاده شده است. در این شبیه سازی از سه جامعه چوله که خلاصه آماری آنها در جدول ۱-۵ ارائه شده است، استفاده گردیده که جامعه اول (rsn) از توزیع چوله نرمال با پارامترهای مکانی ۱۰،

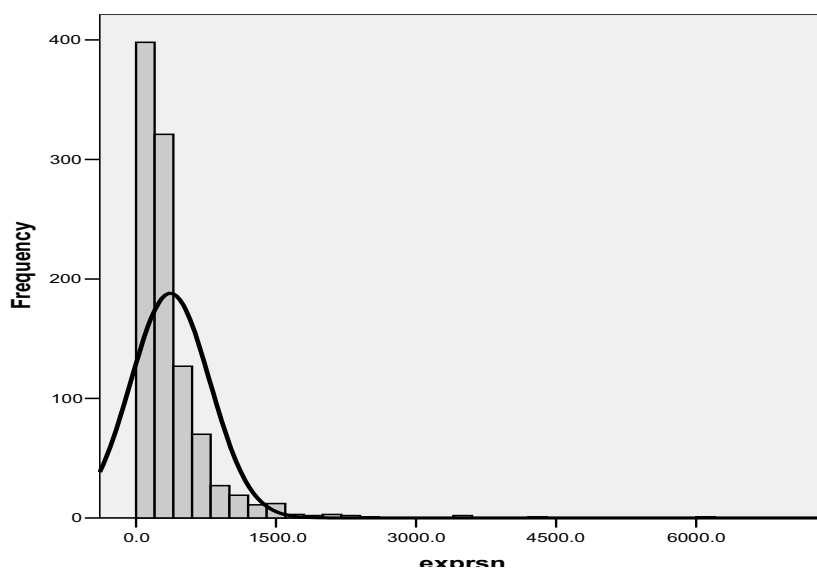
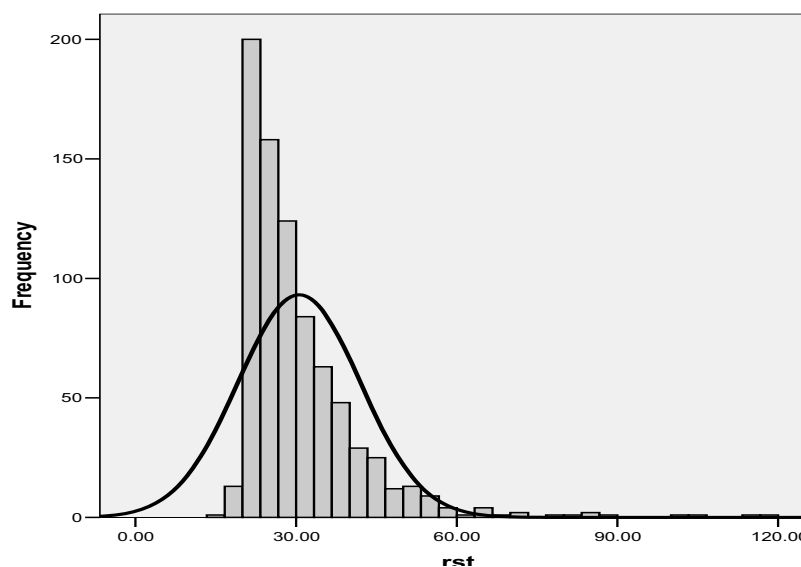
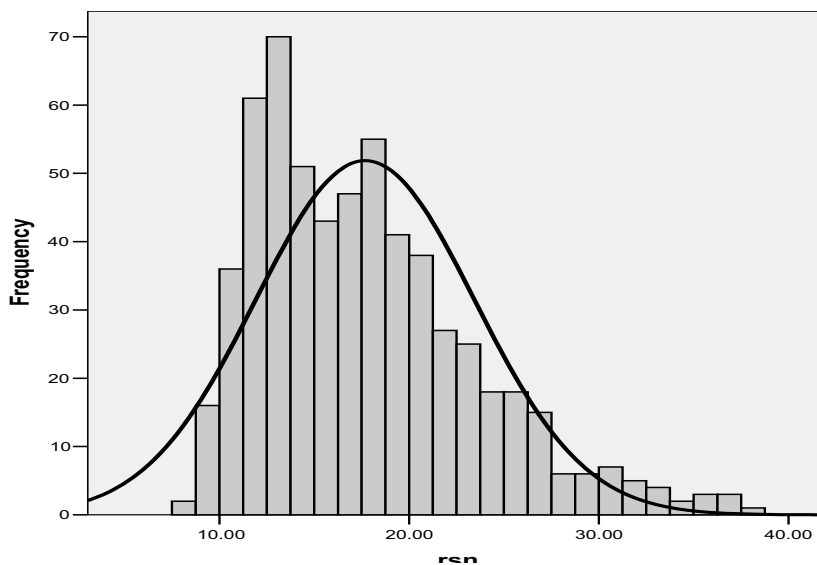
جدول ۱-۵: خلاصه آماری جوامع

جامعه	تعداد (N)	دامنه	میانگین	میانه	واریانس	چولگی
rsn	۶۰۰	۸-۴۰	۱۷/۶۸	۱۶/۷۶	۳۳/۲۸۲	۱/۰۰
rst	۸۰۰	۱۵-۱۲۰	۳۰/۵۷	۲۷/۴۴	۱۳۰/۵۰۶	۳/۰۱
exp(rsn)	۱۰۰۰	۹-۷۰۰۰	۳۶۷/۷۰	۲۴۳/۸۷	۱۸۰۰۲۱/۰	۵/۲۶

$$n_h = \frac{N_h}{\sum_{i=1}^L N_i} \cdot n; \quad h=1, \dots, L,$$

$$n_h = \frac{N_h S_h}{\sum_{i=1}^L N_i S_i} \cdot n; \quad h=1, \dots, L.$$

مقایسات فوق برای ۵ و ۴ و ۳ L انجام شده است. از تخصیص متناسب و تخصیص نیمن برای انتساب نمونه ها به طبقات استفاده شده است که فرمول آنها بصورت روبرو می باشد:



جدول ۵-۲: ساختار طبقات و تخصیص نمونه ها برای جامعه اول به ازای $n=50$

روش	دالینوس - هاج			هندسی			لاوالیه - هایدرولو		
تخصیص	انتساب متناسب								
طبقات	N_h	n_h	CV_h	N_h	n_h	CV_h	N_h	n_h	CV_h
۳	۲۵۸	۲۲	۰/۱۳۲	۲۱۰	۱۸	۰/۱۱۳	۲۷۳	۲۲	۰/۱۴۰
	۲۳۱	۲۰	۰/۰۹۹	۲۹۷	۲۵	۰/۱۳۳	۲۴۰	۲۰	۰/۱۰۷
	۱۱۱	۱۰	۰/۱۴۲	۹۳	۸	۰/۱۳۳	۸۷	۷	۰/۱۳۰
مجموع	۶۰۰	۵۲	۰/۰۱۷	۶۰۰	۵۱	۰/۰۱۸	۶۰۰	۴۹	۰/۰۱۸
۴	۱۹۲	۱۶	۰/۱۱۰	۱۲۴	۱۱	۰/۰۹۱	۲۲۷	۱۷	۰/۱۱۹
	۱۶۹	۱۵	۰/۰۸۲	۲۳۴	۲۰	۰/۱۱۴	۱۹۷	۱۵	۰/۰۸۵
	۱۵۸	۱۴	۰/۰۷۹	۱۹۱	۱۶	۰/۱۰۴	۱۲۶	۱۰	۰/۰۸۰
	۵۱	۷	۰/۱۲۷	۵۱	۵	۰/۱۰۴	۵۰	۴	۰/۱۰۳
مجموع	۶۰۰	۵۲	۰/۰۱۴	۶۰۰	۵۲	۰/۰۱۵	۶۰۰	۴۶	۰/۰۱۴
۵	۱۴۳	۱۲	۰/۰۹۵	۸۰	۷	۰/۰۸۰	۱۸۳	۱۵	۰/۱۰۴
	۱۴۲	۱۲	۰/۰۷۲	۱۸۲	۱۶	۰/۰۸۲	۱۴۶	۱۲	۰/۰۷۳
	۱۵۴	۱۳	۰/۰۶۲	۱۹۶	۱۷	۰/۰۸۲	۱۵۸	۱۳	۰/۰۶۷
	۱۰۶	۹	۰/۱۳۱	۱۱۰	۱۰	۰/۰۸۵	۸۲	۷	۰/۰۶۹
	۵۵	۵	۰/۱۰۸	۳۲	۳	۰/۰۸۲	۳۱	۳	۰/۰۸۰
مجموع	۶۰۰	۵۱	۰/۰۱۱	۶۰۰	۵۳	۰/۰۱۱	۶۰۰	۵۰	۰/۰۱۱
تخصیص	انتساب نیمین								
طبقات	N_h	n_h	CV_h	N_h	n_h	CV_h	N_h	n_h	CV_h
۳	۲۵۸	۱۷	۰/۱۳۵	۲۱۰	۱۱	۰/۱۱۳	۲۵۵	۱۷	۰/۱۳۱
	۲۳۱	۱۷	۰/۰۹۹	۲۹۷	۲۷	۰/۱۳۳	۲۳۲	۱۷	۰/۰۹۹
	۱۱۱	۱۷	۰/۱۴۲	۹۳	۱۳	۰/۱۳۳	۱۱۳	۱۷	۰/۱۴۳
مجموع	۶۰۰	۵۱	۰/۰۱۶	۶۰۰	۵۱	۰/۰۱۷	۶۰۰	۵۱	۰/۰۱۶
۴	۱۹۲	۱۲	۰/۱۰۷	۱۲۴	۶	۰/۰۹۱	۱۹۲	۱۲	۰/۱۱۰
	۱۶۹	۱۲	۰/۰۸۲	۲۳۴	۱۹	۰/۱۱۴	۱۴۷	۹	۰/۰۷۳
	۱۵۸	۱۳	۰/۰۷۹	۱۹۱	۲۰	۰/۱۰۴	۱۷۱	۱۴	۰/۰۷۹
	۸۱	۱۵	۰/۱۲۷	۵۱	۸	۰/۱۰۴	۹۰	۱۷	۰/۱۳۱
مجموع	۶۰۰	۵۲	۰/۰۱۳	۶۰۰	۵۳	۰/۰۱۳	۶۰۰	۵۲	۰/۰۱۳
۵	۱۴۳	۱۰	۰/۰۹۵	۸۰	۴	۰/۰۸۰	۱۱۵	۷	۰/۰۸۹
	۱۴۲	۹	۰/۰۷۲	۱۸۲	۱۲	۰/۰۸۲	۱۴۷	۹	۰/۰۶۹
	۱۵۴	۱۱	۰/۰۶۲	۱۹۶	۱۷	۰/۰۸۲	۱۶۰	۱۲	۰/۰۶۶
	۱۰۶	۱۱	۰/۰۶۸	۱۱۰	۱۳	۰/۰۸۵	۱۱۵	۱۱	۰/۰۶۹

	۵۵	۱۱	۰/۱۰۸	۳۲	۵	۰/۰۸۲	۶۳	۱۳	۰/۱۱۵
مجموع	۶۰۰	۵۲	۰/۰۱۰	۶۰۰	۵۱	۰/۰۱۱	۶۰۰	۵۲	۰/۰۱۰

جدول ۵-۳: ساختار طبقات و تخصیص نمونه ها برای جامعه دوم به ازای $n=۶۰$

روش	دالینوس - هاج			هندسی			لاوالیه - هایدر و گلو		
تخصیص	انتساب متناسب								
طبقات	N_h	n_h	CV_h	N_h	n_h	CV_h	N_h	n_h	CV_h
۳	۴۷۱	۳۶	۰/۱۱۱	۵۰۴	۳۸	۰/۱۲۰	۵۳۷	۴۲	۰/۱۳۰
	۲۴۷	۱۹	۰/۱۰۷	۲۷۹	۲۱	۰/۱۸۰	۲۲۴	۱۸	۰/۱۳۲
	۸۲	۷	۰/۲۸۷	۱۷	۲	۰/۲۲۲	۳۹	۳	۰/۲۷۲
مجموع	۸۰۰	۶۲	۰/۰۲۳	۸۰۰	۶۱	۰/۰۲۱	۸۰۰	۶۳	۰/۰۲۲
۴	۳۳۲	۲۵	۰/۰۷۷	۳۲۳	۲۵	۰/۰۷۵	۴۷۹	۳۰	۰/۱۱۳
	۲۵۵	۲۰	۰/۰۷۷	۳۹۳	۲۹	۰/۱۴۳	۲۳۷	۱۸	۰/۰۹۰
	۱۵۶	۱۲	۰/۰۹۴	۷۵	۶	۰/۱۳۴	۷۵	۱۰	۰/۱۴۰
	۵۷	۵	۰/۲۸۲	۹	۲	۰/۱۵۱	۹	۲	۰/۱۵۵
مجموع	۸۰۰	۶۲	۰/۰۱۹	۸۰۰	۶۲	۰/۰۱۷	۸۰۰	۶۰	۰/۰۱۸
۵	۲۲۶	۱۷	۰/۰۵۷	۱۸۲	۱۴	۰/۰۵۲	۳۷۲	۲۷	۰/۰۸۵
	۲۶۵	۲۰	۰/۰۶۸	۴۲۸	۳۲	۰/۱۲۰	۲۴۳	۱۸	۰/۰۷۵
	۱۸۰	۱۴	۰/۰۷۱	۱۵۳	۱۲	۰/۱۱۲	۱۳۳	۱۰	۰/۰۸۹
	۸۹	۷	۰/۰۸۰	۲۹	۳	۰/۱۱۱	۴۳	۴	۰/۱۱۰
	۴۰	۳	۰/۲۷۳	۸	۲	۰/۱۳۹	۹	۲	۰/۱۵۱
مجموع	۸۰۰	۶۱	۰/۰۱۸	۸۰۰	۶۳	۰/۰۱۴	۸۰۰	۶۱	۰/۰۱۴
تخصیص	انتساب نینم								
طبقات	N_h	n_h	CV_h	N_h	n_h	CV_h	N_h	n_h	CV_h
۳	۴۷۱	۲۲	۰/۱۱۱	۵۰۴	۲۴	۰/۱۲۰	۴۲۴	۱۷	۰/۰۹۸
	۲۴۷	۱۶	۰/۱۰۷	۲۷۹	۳۲	۰/۱۸۰	۲۸۹	۱۹	۰/۱۲۰
	۸۲	۲۳	۰/۲۸۷	۱۷	۵	۰/۲۲۳	۸۷	۲۳	۰/۲۸۸
مجموع	۸۰۰	۶۱	۰/۰۱۷	۸۰۰	۶۱	۰/۰۱۸	۸۰۰	۵۹	۰/۰۱۷
۴	۳۳۲	۱۳	۰/۰۷۷	۳۲۳	۱۱	۰/۰۷۵	۳۵۱	۱۵	۰/۰۸۰
	۲۵۵	۱۳	۰/۰۷۷	۳۹۳	۳۷	۰/۱۴۳	۲۴۲	۱۲	۰/۰۷۴
	۱۵۶	۱۳	۰/۰۹۴	۷۵	۱۱	۰/۱۳۴	۱۶۰	۱۵	۰/۱۱۲
	۵۷	۲۲	۰/۲۸۲	۹	۳	۰/۱۵۱	۴۷	۱۹	۰/۲۸۷
مجموع	۸۰۰	۶۱	۰/۰۱۳	۸۰۰	۶۲	۰/۰۱۴	۸۰۰	۶۱	۰/۰۱۳
۵	۲۲۶	۸	۰/۰۵۷	۱۸۲	۵	۰/۰۵۲	۲۵۳	۱۰	۰/۰۶۱

	۲۶۵	۱۳	۰/۰۶۸	۴۲۸	۳۳	۰/۱۲۰	۲۲۹	۱۰	۰/۰۵۹
	۱۸۰	۱۲	۰/۰۷۱	۱۵۳	۱۶	۰/۱۱۲	۱۷۱	۱۱	۰/۰۶۷
	۸۹	۹	۰/۰۸۰	۲۹	۵	۰/۱۱۱	۱۰۸	۱۲	۰/۱۰۰
	۴۰	۲۰	۰/۲۷۳	۸	۳	۰/۱۳۹	۳۹	۱۹	۰/۲۷۲
مجموع	۸۰۰	۶۲	۰/۰۱۰	۸۰۰	۶۲	۰/۰۱۱	۸۰۰	۶۲	۰/۰۱۰

جدول ۴-۵: ساختار طبقات و تخصیص نمونه ها برای جامعه سوم به ازای $n=100$

روش	دالینوس - هاج			هندسی			لاوالیه - هایدر و گلو		
تخصیص	انتساب متناسب								
طبقات	N_h	n_h	CV_h	N_h	n_h	CV_h	N_h	n_h	CV_h
۳	۴۴۹	۴۵	۰/۳۷۱	۱۹۰	۱۹	۰/۲۸۹	۷۶۲	۵۶	۰/۵۱۳
	۴۰۶	۴۱	۰/۲۹۵	۷۳۴	۷۴	۰/۵۳۵	۲۰۷	۲۵	۰/۳۰۰
	۱۴۵	۱۵	۰/۶۳۶	۷۶	۸	۰/۵۶۵	۳۱	۴	۰/۴۰۵
مجموع	۱۰۰۰	۱۰۱	۰/۰۷۰	۱۰۰۰	۱۰۱	۰/۰۶۹	۱۰۰۰	۸۵	۰/۰۶۳
۴	۳۲۴	۳۳	۰/۳۳۰	۶۵	۷	۰/۲۲۵	۶۶۰	۴۰	۰/۴۵۹
	۴۵۱	۴۶	۰/۲۹۱	۵۵۷	۵۶	۰/۳۶۴	۲۷۰	۲۵	۰/۲۷۳
	۱۷۷	۱۸	۰/۲۳۸	۳۵۱	۳۶	۰/۴۱۴	۶۲	۱۳	۰/۲۸۳
	۴۸	۵	۰/۵۲۸	۲۷	۳	۰/۴۸۸	۸	۲	۰/۲۳۴
مجموع	۱۰۰۰	۱۰۲	۰/۰۵۶	۱۰۰۰	۱۰۲	۰/۰۵۷	۱۰۰۰	۸۰	۰/۰۵۰
۵	۲۴۰	۲۴	۰/۳۰۱	۲۶	۳	۰/۲۰۷	۵۲۲	۳۴	۰/۳۹۴
	۲۸۶	۲۹	۰/۱۷۲	۳۱۷	۳۲	۰/۲۸۲	۳۲۵	۲۵	۰/۲۴۰
	۲۷۶	۲۸	۰/۱۹۲	۴۹۴	۵۰	۰/۳۳۵	۱۱۶	۱۲	۰/۱۹۰
	۱۳۸	۱۴	۰/۱۸۲	۱۵۲	۱۱	۰/۳۳۶	۳۳	۷	۰/۲۳۱
	۶۰	۶	۰/۵۴۵	۱۱	۲	۰/۴۱۰	۴	۲	۰/۲۳۴
مجموع	۱۰۰۰	۱۰۱	۰/۰۵۷	۱۰۰۰	۱۰۳	۰/۰۴۳	۱۰۰۰	۸۰	۰/۰۴۱
تخصیص	انتساب نیمین								
طبقات	N_h	n_h	CV_h	N_h	n_h	CV_h	N_h	n_h	CV_h
۳	۴۴۹	۱۴	۰/۳۷۱	۱۹۰	۳	۰/۲۸۹	۶۲۲	۳۰	۰/۴۳۶
	۴۰۶	۲۷	۰/۲۹۵	۷۳۴	۶۶	۰/۵۳۵	۳۰۶	۲۷	۰/۲۵۵
	۱۴۵	۶۱	۰/۶۳۶	۷۶	۳۲	۰/۵۶۵	۷۲	۳۳	۰/۵۶۰
مجموع	۱۰۰۰	۱۰۲	۰/۰۳۸	۱۰۰۰	۱۰۱	۰/۰۴۰	۱۰۰۰	۹۰	۰/۰۴۰
۴	۳۲۴	۱۰	۰/۳۳	۶۵	۲	۰/۲۲۵	۴۹۸	۲۲	۰/۳۹۰
	۴۵۱	۳۱	۰/۲۹۱	۵۵۷	۲۵	۰/۳۶۴	۳۱۱	۱۹	۰/۲۰۵
	۱۷۷	۲۴	۰/۲۳۸	۳۵۱	۵۶	۰/۴۱۴	۱۵۳	۲۱	۰/۲۵۰

	۴۸	۳۷	۰/۵۲۸	۲۷	۱۹	۰/۴۸۸	۳۸	۲۹	۰/۵۱۵
مجموع	۱۰۰۰	۱۰۲	۰/۰۲۷	۱۰۰۰	۱۰۲	۰/۰۳۰	۱۰۰۰	۹۱	۰/۰۳۰
۵	۲۴۰	۷	۰/۳۰۱	۲۶	۲	۰/۲۰۷	۳۳۹	۱۳	۰/۳۳۴
	۲۸۶	۱۰	۰/۱۷۲	۳۱۷	۹	۰/۲۸۲	۲۸۴	۱۲	۰/۱۷۲
	۲۷۶	۱۹	۰/۱۹۲	۴۹۴	۴۲	۰/۳۳۵	۲۱۴	۱۵	۰/۱۶۲
	۱۳۸	۱۷	۰/۱۸۲	۱۵۲	۳۸	۰/۳۳۶	۱۲۵	۲۰	۰/۲۱۰
	۶۰	۵۰	۰/۵۴۵	۱۱	۱۱	۰/۳۹۹	۳۸	۳۶	۰/۵۱۰
مجموع	۱۰۰۰	۱۰۳	۰/۰۲۱	۱۰۰۰	۱۰۲	۰/۰۲۲	۱۰۰۰	۹۶	۰/۰۲۱

Annals of Mathematical Statistics, 30, 219-229.

- [4]. Gunning, P. and Horgan, J. (2004). A new algorithm for the construction of stratum boundaries in skewed populations. *Survey Methodology*, 30, 159-166.
- [5]. Gunning, P., Horgan, J. M. and Yancy, W. (2004). Geometric stratification of accounting data. *de Contaduria y Administration*, 214.
- [6]. Gunning, P., Horgan, J. M. and Koegh, G. (2005). Efficient Pareto stratification, *The Mathematical Proceedings of the Royal Irish Academy*.
- [7]. Horgan, J. M. (2006). Stratification of skewed populations: A review. *International Statistical Review*, 74(1): 67-76.
- [8]. Lavellée, P. and Hidiroglou, M. (1988). On the stratification of skewed populations. *Survey Methodology*, 14, 33-43.

۶- نتیجه گیری

با توجه به آنچه در بالا بیان شد، روش هندسی همانند روش دالینوس - هاج یک روش تقریبی است و بطور کلی یک روش بهینه بشمار نمی آید. اما این روش دارای الگوریتمی است که به آسانی قابل اجرا بوده و بر خلاف روش لاولیه - هایدرولوگو دارای الگوریتم تکرار چند مرحله ای نیست. نتایج بدست آمده از جداول بالا نشان می دهد که در دو جامعه اول که دارای چولگی کمتری هستند تفاوت چندانی میان روش های مورد بررسی، وجود ندارد. اما روش هندسی به سبب محاسبات کمتر به دو روش دیگر برتری دارد. در جامعه سوم که دارای چولگی بیشتری است روش بهینه لاولیه - هایدرولوگو بطور قابل ملاحظه ای نسبت به دو الگوریتم تقریبی روش هندسی و دالینوس - هاج بهتر عمل می کند. البته در انتساب نیمین کمی این برتری تعدیل شده اما همچنان برتری نسبی با روش بهینه لاولیه - هایدرولوگو است. با توجه به نتایج به دست آمده توصیه می شود تا از روش هندسی در جوامعی استفاده شود که دارای چولگی کمتر از چهار هستند.

منابع

- [1]. Cochran, W. (1961). Comparison of methods for determining stratum boundaries. *Bulletin of the International Statistical Institute*, 38, 2.
- [2]. Dalenius, T. and Hodges, J. (1957). The choice of stratification points. *Skandinavisk Aktuarietidskrift*, 198-203.
- [3]. Ekman, G. (1959). An approximation useful in univariate stratification. *The*