

نمونه گیری خوشه ای سازوار بر اساس آماره های ترتیبی

فرخنده السادات سجادی^۱ محمد صالحی^۲

چکیده

یکی از طرحهایی که برای نمونه گیری از جوامع خوشه ای و کمیاب به کار می رود، نمونه گیری خوشه ای سازوار است. در این طرح بعد از انتخاب یک نمونه اولیه، واحدهای دیگر جامعه در صورتی به نمونه اضافه می شوند که در شرطی صدق کنند. در بسیاری از موارد، تعیین این شرط قبل از نمونه گیری مشکل است. در این صورت می توان از اطلاعات به دست آمده طی نمونه گیری استفاده کرد. نمونه گیری خوشه ای سازوار بر اساس آماره های ترتیبی [۲]، امکان استفاده از این اطلاعات را به ما می دهد. در این مقاله با استفاده از شبیه سازی، کارایی این طرح نمونه گیری مورد بررسی قرار می گیرد. هرچه حالت خوشه ای جامعه ها افزایش پیدا کند، کارایی این روش زیاد می شود. همچنین تأثیر اندازه واحدهای اولیه نمونه گیری در میزان کارایی نیز، بررسی می شود.

واژه های کلیدی: نمونه گیری شبکه ای، واحد مرزی، شبیه سازی، قاعدة توقف، نمونه گیری با احتمال نابرابر.

۱. مقدمه

علاقة مشاهده شده در طول مطالعه دارد. یکی از اهداف نمونه گیری سازوار این است که از مشخصه های مفید جامعه مورد مطالعه استفاده کرده تا برآورده دقیقت برای پارامترهای جامعه، مانند چگالی و یا فراوانی جامعه، نسبت به دیگر طرحهای متداول نمونه گیری به دست آورد.

۲. نمونه گیری خوشه ای سازوار

این روش نمونه گیری که در سال ۱۹۹۰ به وسیله تامسون^۳ [۱] مطرح شد، یکی از روشهای مفید نمونه گیری برای برآورد پارامتر در جامعه های خوشه ای و کمیاب است. در این روش با استفاده از بعضی روشهای نمونه گیری متداول مانند نمونه گیری تصادفی ساده

معمولأ بیشتر توجه در نظریه نمونه گیری و روش تحقیق، به طرحهای نمونه گیری است که روشهای انتخاب به مشاهداتی که در طول مطالعه به دست می آیند، بستگی نداشته باشد. بنابراین می توان نمونه کامل واحدها را قبل از عملیات میدانی مشخص کرد. اما در بسیاری از وضعیتهاي نمونه گیری، مایلیم تصمیمهای خود را در طول مطالعه، بر اساس آنچه تاکنون مشاهده کرده ایم، اتخاذ کنیم.

نمونه گیری سازوار، نمونه گیری است که در آن انتخاب واحدهایی که در نمونه قرار می گیرند، بستگی به مقادیر متغیر مورد

^۱ دانشکده علوم ریاضی، دانشگاه صنعتی اصفهان

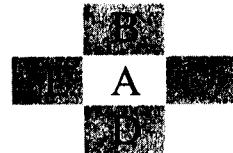
^۲ دانشجوی کارشناسی ارشد دانشکده علوم ریاضی، دانشگاه صنعتی اصفهان

مشاهده شده و به نمونه اضافه می شوند. بعضی از این واحدهای اضافه شده، ممکن است در شرط C صدق کنند (که در این صورت همسایگان این واحدها نیز به نمونه اضافه می شوند) و برخی دیگر در این شرط C صدق نمی کنند. به همین ترتیب برای واحدهایی که باز در شرط C صدق کنند، همسایگان آنها به نمونه اضافه می شوند و این کار ادامه می یابد تا اینکه خوشه ای از واحدها به دست آیند. مرز این خوشه را واحدهایی تشکیل می دهند که واحد مرزی نامیده می شوند. این واحدها در شرط مورد نظر C صدق نکرده اند. یک خوشه را بدون در نظر گرفتن واحدهای مرزی آن شبکه می نامیم. واحدی که در شرط C صدق نکند ولی در نمونه اولیه انتخاب شده باشد، تشکیل یک شبکه با اندازه ۱ را می دهد. در این طرح نمونه گیری ممکن است تعدادی قابل توجه از واحدهای نمونه اولیه در شرط از پیش مشخص شده C صدق کنند، به طوری که به دلیل محدودیتهای مالی یا زمانی قادر به اتمام طرح نمونه گیری نباشیم. برای حل این مشکل راه حلهای مختلفی ارائه شده است. یکی از آنها که مورد توجه محققان دریابی قرار گرفته است، نمونه گیری خوشه ای سازوار بر اساس آماره های ترتیبی است که در بخش بعد معرفی می شود. شکل ۲، نمونه گیری سازوار خوشه ای برای برآورد تعداد اشیاء نقطه ای در ناحیه مورد مطالعه با ۶۴ واحد را نشان می دهد. یک نمونه تصادفی اولیه با ۴ واحد (واحدهای خاکستری) در شکل نشان داده شده است. واحدهای همسایه به نمونه اضافه می شوند، هرگاه حداقل یک شیء از جامعه در واحد انتخاب شده مشاهده شود. نمونه نهایی شامل واحدهای خاکستری و واحدهای هاشور زده است.

۳. نمونه گیری سازوار خوشه ای بر اساس آماره های ترتیبی

در نمونه گیری سازوار خوشه ای، هرگاه مقدار متغیر مورد مطالعه برای یک واحد در نمونه در شرطی معین صدق کند، واحدهای همسایه آن به نمونه اضافه می شوند. در بیشتر کاربردهای عملی نمونه گیری سازوار، شرط برای اضافه کردن همسایه ها به

(بدون جایگذاری - با جایگذاری)، نمونه گیری سیستماتیک و ... یک نمونه اولیه از واحدهای جامعه انتخاب می شود، و هرگاه متغیر مورد مطالعه واحد انتخاب شده در شرطی صدق کند، واحدهایی که در همسایگی آن واحد قرار دارند، به نمونه اضافه می شوند. همسایگی می تواند به مفهوم عادی آن تعریف شود، تنها با این شرط که این رابطه همسایگی، یک رابطه متقاضان است. بدین معنا که اگر واحد A در همسایگی واحد Z باشد، آنگاه واحد Z نیز در همسایگی واحد A است. این همسایگیها به مقادیر متغیر مورد مطالعه جامعه بستگی ندارند. همچنین این همسایگیها ممکن است بر اساس یک رابطه خاص که بین واحدها وجود دارد، تعریف شوند. همسایگی می تواند به صورت یک منطقه جغرافیایی نیز در نظر گرفته شود. در شکل ۱، همسایه های واحد A، واحدهای B، C و D هستند که به ترتیب در بالا، سمت راست، پایین و سمت چپ واحد A قرار دارند.



شکل ۱

شرطی که برای انتخاب نمونه های اضافی بعد از نمونه گیری اولیه وجود دارد، (اضافه کردن واحدهای همسایه واحد نمونه اولیه)، به وسیله یک فاصله یا یک مجموعه C در برداشت مطالعه معمین می شود.

فرض می کنیم از یک جامعه متناهی با اندازه N، نمونه گیری می کیم و اعضا را با برجسبهای N, ..., ۲, ۱ مشخص کرده و مقادیر مورد علاقه را به صورت $\{y_N, y_{N-1}, \dots, y_1\} = y$ نشان می دهیم، به طوری که y_i مقدار متغیر مورد مطالعه برای واحد i است. در این صورت می گوییم واحد A در شرط مورد نظر که آن را با C نمایش می دهیم، صدق می کند، اگر $y_i \in C$ باشد.

بعد از نمونه گیری اولیه، هنگامی که یک واحد انتخاب شده در شرط C صدق کرد، همه واحدهایی که در همسایگی آن هستند

استفاده کنندگان این طرح، از قبل، تعداد مراحلی را که به نمونه سازوار اضافه می‌شود مشخص می‌کنند و بعد از آن نمونه گیری را متوقف می‌کنند. قاعدة توقف S را به عنوان سطح S ام نمونه گیری سازوار - جدا از واحد انتخاب شده در نمونه گیری اولیه - تعریف می‌کنیم که بعد از این سطح، فرایند نمونه گیری متوقف می‌شود.

		S_r	S_r	S_r		
	S_r	S_r	S_r	S_r	S_r	
S_r	S_r	S_r	O	S_r	S_r	
		S_r	S_r	S_r		S_r
			S_r	S_r		

شکل ۳

شکل ۳ نمونه گیری خوش‌ای سازوار با قاعدة توقف $S = 3$ است. (O نشان دهنده واحد نمونه اولیه و واحدهای سایه زده شده، نشان دهنده واحدهای مرزی اند).

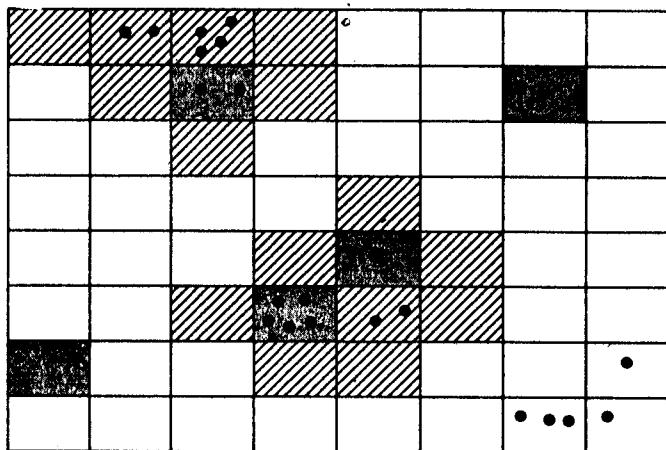
۴. برآوردهای هنسن - هورویتس^۴ و هورویتس - تامسون^۵

یک برآوردهگر نااریب برای میانگین، برآوردهگر هنسن - هورویتس است که برای نمونه گیری خوش‌ای سازوار بر اساس آماره‌های ترتیبی به صورت زیر است:

$$\hat{\mu}_{HH} = \frac{1}{n} \sum_{i=1}^n w_i = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{x_i}$$

Hansen - Hurwitz⁴

Horvitz - Thompson⁵



شکل ۲

نمونه، به صورت $C > y_i$ است که C مقدار ثابتی است. اگر C مقدار نامناسب کوچکی انتخاب شود، ممکن است تعداد واحدهای لازم برای اضافه شدن به مجموعه نمونه خیلی زیاد باشد. به طوری که قادر به اتمام طرح نمونه گیری نباشیم. همچنین انتخاب مقدار نامناسب و بزرگی برای C نیز ممکن است باعث شود تا هیچ یک از واحدهای اولیه در شرط مورد نظر صدق نکند و عملانه نمونه نهایی شامل همان نمونه اولیه باشد. در این گونه موارد تامسون در سال ۱۹۹۶ طرح نمونه گیری سازوار خوش‌ای بر اساس آماره‌های ترتیبی را مطرح کرد [۲]. بدین ترتیب که فرض کنید در نمونه گیری اولیه با اندازه n ، مقادیر y_1, y_2, \dots, y_n را برابر واحدهای $1, 2, \dots, n$ مشاهده کنیم. آنها را به صورت $y_1 \leq y_2 \leq \dots \leq y_n$ مرتب می‌کنیم به طوری که y_i ، مقدار i امین آماره ترتیبی باشد. برای هر واحد i در نمونه، همسایه‌های آن، به نمونه اضافه می‌شوند، اگر y_i مربوط به آن واحد از مقدار $(n - i)$ امین آماره ترتیبی بزرگتر باشد.

استفاده از نمونه گیری سازوار بر اساس آماره‌های ترتیبی، کنترل قابل توجهی روی حجم نمونه نهایی دارد. اما اگر با یک جامعه بسیار غیر طبیعی با خوش‌های بزرگ رو به رو شویم، بعد از انتخاب نمونه اولیه، باز هم مجبور به انتخاب واحدهای زیادی هستیم که ممکن است برای ما عملی نباشد. به همین دلیل بعضی از

ماهی لوتوی اقیانوس آرام (POP)^۷ یکی از مهمترین گونه های تجاری است. صید این گونه ماهی در سال ۱۹۶۵ به بالاترین مقدار خود (حدود ۳۵۰۰۰۰ تن) در دهه ۱۹۶۰ رسید. صید زیاد موجب کاهش نژاد این گونه ماهی شد. بررسیهای ماهیگیری در سرتاسر خلیج، هر سه سال یک بار، از سال ۱۹۸۴، برای برآورد بیومس^۸ ماهیهای صخره ای انجام شده است [۳]. در سال ۱۹۹۸ یک پروژه مشترک برای یافتن برآوردهای بهتر از فراوانی ماهیهای صخره ای بین دانشگاه فربانکس آلاسکا^۹، صنعت شیلات و آزمایشگاه ملی آک بی^{۱۰}، در ناحیه کوچکی از خلیج آلاسکا آغاز شد [۴]. فرضیه اساسی مسؤولان طرح این بود که نمونه گیری سازوار در دستیابی به برآوردهای دقیق از فراوانی ماهیهای صخره ای، کاراتر از نمونه گیری تصادفی طبقه بندی استاندارد است. فرضیه بعدی آنها عبارت بود از این که برآورد فراوانی POP از طبق نمونه گیری خوش ای سازوار، سودمندتر از برآورد گونه های Shortraker و Rougheye از ماهیهای صخره ای است. به عبارت دیگر استفاده از نمونه گیری خوش ای سازوار برای جمعیت‌هایی مانند Shortraker و Rougheye که خوش ای نیستند، مفید نخواهد بود، زیرا آنها معتقد بودند که POP ها در توزیعشان نسبت به Shortraker خوش ای تر هستند. یک مطالعه دو هفته ای در اوت ۱۹۹۸ بر طبق طرح نمونه گیری خوش ای سازوار در ۶ طبقه، نزدیک پورت لاک بانک^{۱۱} در خلیج آلاسکا انجام گرفت (چهار طبقه برای POP و دو طبقه برای Shortraker\Rougheye). این طبقه بندی بر اساس انواع زیستگاههای متفاوت و نواحی فرض شده از چگالیهای مختلف بود. در هر طبقه بندی در ۱۲ یا ۱۵ ایستگاه تصادفی که توسط رایانه مشخص شده بود، نمونه گیری اولیه اجرا

که y_i و y_j به ترتیب میانگین و مجموع کل مشاهده در شبکه ای است که با واحد α_k نمونه اولیه اشتراک دارد. برآورد گر هوروپیس - تامسون برای میانگین، به صورت زیر است:

$$\hat{\mu}_{HT} = \frac{1}{N} \sum_{k=1}^K \frac{y_k^*}{\alpha_k}$$

که y_k^* مجموع مقادیر y برای شبکه k ام، K تعداد شبکه های مجزا در نمونه، و α_k احتمال این است که شبکه k در نمونه باشد. اگر

x_k واحد، در شبکه k ام وجود داشته باشد، آنگاه

$$\alpha_k = 1 - \frac{\binom{N - x_k}{n}}{\binom{N}{n}}$$

در نمونه گیری خوش ای سازوار بر اساس آماره های ترتیبی، برخلاف نمونه گیری خوش ای سازوار، برآورد گر هوروپیس - تامسون اریب است.

۵. مثالی از کاربرد نمونه گیری سازوار بر اساس آماره های ترتیبی و اعمال قاعدة توقف

بررسیهای ماهیگیری (طی کردن مسیر مورد نظر در دریا با کشتنی و شمارش ماهیها با صید کردن آنها و یا استفاده از روشهای مختلف دیگر) با طرح نمونه گیری ساده طبقه بندی، برای ارزیابی چگونگی جمعیت ماهیها، از رایجترین طرحهایی است که به کار می رود. اگر چه این روش نمونه گیری برای گونه هایی از ماهی، کارایی خوبی دارد اما برای گونه هایی که توزیع جمعیتشان به صورت خوش ای است (در مکانهایی خاص تجمع کرده اند)، ممکن است طرحی کارا نباشد. ماهیهای صخره ای^{۱۲} در شمال شرقی اقیانوس آرام، نمونه ای از این گونه ها هستند. مجموعه ماهیهای صخره ای در خلیج آلاسکا از گونه های مختلف تشکیل می شود.

Pacific Ocean Perch^۷

BIOMASS^۸

University of Alaska Fairbanks^۹

Auke Bay^{۱۰}

Portlock Bank^{۱۱}

Rockfish^{۱۲}

ماهیهای لوتوی اقیانوس آرام بهتر از نمونه گیری تصادفی ساده در طبقات مختلف عمل کرده است. در مقایسه با نمونه گیری تصادفی ساده، مقدار SE برای برآورد گرهای سازوار از فراوانی ماهیهای لوتوی اقیانوس آرام در هر ۴ طبقه کاهش یافته، اما نمونه گیری سازوار در مقایسه با نمونه گیری تصادفی ساده برای ماهیهای Rougheye و Shortraker در طبقه شمال که توزیع جمعیت آنها به صورت خوشه ای نبوده، خوب عمل نکرده است. برای طبقه جنوبی، نتایج حاصل از نمونه گیری تصادفی ساده و نمونه گیری سازوار تقریباً یکسان است. بنابراین با اجرای نمونه گیری سازوار سود بیشتری نسبت به استفاده از نمونه گیری تصادفی ساده در برآورد فراوانیها به دست نخواهیم آورد. پاسخ به این سؤال که «آیا استفاده از نمونه گیری سازوار سودی را برای ماهیهای Rougheye و Shortraker به دست می دهد یا نه؟»، منوط به انجام تحقیقات بیشتر است.

نتایج بالا که از زبان متخصصان علوم دریایی بازگو شده است، اگرچه چندان پایه آماری قوی و مستدلی ندارند، ولی برای آشنایی خواننده با نوع استنتاجی که توسط دانشمندان علوم دیگر صورت می گیرد، بیان شده اند. برای بررسی خواص طرح نمونه گیری خوشه ای سازوار بر اساس آماره های ترتیبی با پایه های قویتر، شبیه سازی کوچکی در قسمت بعد آورده می شود.

۶. مقایسه کارایی نمونه گیری تصادفی ساده و نمونه گیری خوشه ای سازوار بر اساس آماره های ترتیبی

فرض کنید ناحیه مورد مطالعه را به ۴۰۰ واحد مربع تقسیم کرده ایم که گونه خاص از نوعی گیاه، تنها در بعضی از نقاط آن می روید. اگر گیاهی که تکثیر می یابد و گیاهان دیگری از نوع خود را در اطرافش تولید می کند، گیاه مولد، و گیاهان تولید مثل شده را گیاهان فرعی بنامیم، جمعیتی به شرح زیر را بر این اساس تولید می کنیم:

شد. بعد از اتمام نمونه گیری اولیه، نمونه گیری خوشه ای سازوار بر اساس آماره های ترتیبی و اعمال قاعدة توقف اجرا شد. بدین ترتیب که اگر y CPUE ^{۱۲} گونه هدف تعریف کنیم و مقادیر مشاهده شده y از نمونه گیری اولیه را به صورت

$$y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)}$$

که ۱۱، برابر با ۱۵ یا ۱۲ است، نشان دهیم، همسایگان واحدهای اولیه در صورتی به نمونه اضافه می شوند که مقدار CPUE مشاهده شده در آنها بزرگتر از $y_{(n-1)}$ باشد. برای ناحیه مورد مطالعه CPUE، مقدار Shortraker-Rougheye ترکیب شد. در ابتدا مقدار ۲ برابر با ۳ انتخاب شد.

در طبقات slop_Gully، East Gully و West Gully Intersection جنوب - شرق - غرب واحد انتخاب شده در نمونه اولیه و در طبقات slop و Shortraker_Rougheye تعیین شده، همسایگی به صورت خطی تعریف شد.

برای محدود کردن نمونه گیری سازوار، یک قاعدة توقف دلخواه اعمال شد. برای طبقاتی که نمونه گیری سازوار در شمال، جنوب، شرق و غرب واحد اولیه اجرا شد، انتخاب مقدار ۳ برای S موجب شد تا حداقل ۲۴ واحد سازوار اطراف هر واحدی که در شرط مورد نظر صدق می کرد، بررسی شود. در طبقاتی که نمونه گیری سازوار به صورت خطی انجام می شد، با انتخاب مقدار ۶ برای S، حداقل ۱۲ واحد سازوار مورد بررسی قرار گرفتند.

جدول ۱ برآوردهای به دست آمده برای فراوانی ماهیهای صخره ای و انحراف استاندارد (SE) این برآورد گرهای را نشان می دهد. نشان دهنده برآورد به دست آمده از نمونه گیری تصادفی ساده است. HH و HT به ترتیب بیانگر مقادیر برآورد گرهای سازوار هنسن - هورویتس و هورویتس - تامسون هستند. در این جدول مشاهده می شود، نمونه گیری سازوار برای

مقدادر $\hat{\mu}_{HH}$ ، $\hat{\mu}_{HT}$ و اندازه نمونه مؤثر که با $E(v')$ نشان داده می شود و

$$E(v') = \sum_{k=1}^N \alpha_k$$

محاسبه می شوند. واریانس میانگین در نمونه گیری تصادفی ساده را نیز از رابطه زیر به دست می آوریم:

$$\text{Var}(\hat{\mu}_{\text{sss}}) = (1 - E(v')/N)\sigma^2/E(v')$$

سرانجام مقدادر ($R = 1000$ ، $\hat{\mu}_i = \hat{\mu}_{HH}$ or $\hat{\mu}_{HT}$)

$$\bar{\mu} = \sum_{i=1}^R \frac{\hat{\mu}_i}{R}, \quad \text{VARP} = \sum_{i=1}^R (\hat{\mu}_i - \bar{\mu})^2 / R$$

$$\text{MSE}(\hat{\mu}) = (\bar{\mu} - \mu)^2 + \text{VARP}$$

$$\text{RE}(\hat{\mu}) = \text{Var}(\hat{\mu}_{\text{sss}}) / \text{MSE}(\hat{\mu})$$

را برای هر برآوردگر سازوار هنسن - هورویتس یا هورویتس - تامسون به دست می آوریم. نمودارهای شکل ۴ کارایی برآوردگرهای سازوار را نسبت به برآوردگر میانگین در نمونه گیری تصادفی ساده (بدون جایگذاری) نشان می دهد.

۷. بحث و نتیجه گیری

مثالی که در مورد نمونه گیری از جمعیت ماهیهای صخره ای بیان شد، از مقاله ای تحت عنوان «نمونه گیری خوش ای سازوار جمعیت ماهیهای صخره ای» گرفته شده است [۴]. ما سعی کرده ایم تا آنجا که ممکن است، مقاله حاضر را از فرمول و نظریه موضوع مورد مطالعه دور نگه داریم، به این امید که مجموعه بزرگتری از خوانندگان این مجله را به خواندن مقاله ترغیب کنیم. به طور سنتی، جامعه دانشگاهی آماری ایران، توجه خاصی به مبانی نظری و ریاضیات داشته است و این امر مایه مباحثات و افتخار است. ولی امروزه که علم آمار توسط طیف وسیعی از محققان و کاربران علوم دیگر مورد استفاده قرار می گیرد، موجب شده تا وظيفة آماردانان

تعدادی گیاه مولد از توزیع پواسون با میانگین λ تولید کرده و با استفاده از توزیع یکوتاخت در فاصله $(0, 20)$ مکان این گیاهان را در ناحیه مورد نظر مشخص می کنیم. گیاهان فرعی هر گیاه مولد را ب طبق یک فرایند پواسون با میانگین λ تولید کرده و با استفاده از توزیع نرمال دو متغیره ای که میانگین آن مختصات مکان گیاه مولد در ناحیه و واریانس آن $I_{xx} = 5$ است، در اطراف گیاه مولد پراکنده می کنیم. با مقدادر کوچک λ خوشه هایی از گیاه را به دست می آوریم که ناحیه کوچکی را می پوشانند. مقدادر بزرگ λ خوشه هایی با پراکندگی زیاد در ناحیه بزرگی را نتیجه می دهد. بدین معنی که در هر مربع تعداد گیاه کم، ولی مربعهای زیادی دارای این پوشش گیاهی هستند. هر یک از واحدهای اولیه را به صورت، مربع واحد در نظر می گیریم. برای مطالعه چگونگی در نظر گرفتن اندازه واحدهای اولیه، ناحیه را به سه حالت زیر مورد بررسی قرار می دهیم:

- ۱) ناحیه، مربعی به طول ضلع ۱۰ واحد مربع است ($N = 10 \times 10$).
- ۲) ناحیه، مستطیلی به طول ۲۰ و عرض ۱۰ واحد مربع است ($N = 20 \times 10$).
- ۳) ناحیه، مستطیلی به طول ۱۰ و عرض ۲۰ واحد مربع است ($N = 10 \times 20$).

میانگین و واریانس سه جمعیت تولید شده عبارت اند از:

$$N = 10 \times 10, \mu = 5 / 80, \sigma^2 = 762 / 6338$$

$$N = 20 \times 10, \mu = 2 / 920, \sigma^2 = 258 / 4617$$

$$N = 10 \times 20, \mu = 2 / 920, \sigma^2 = 204 / 6928$$

برای مقایسه کارایی نمونه گیری تصادفی ساده و نمونه گیری خوش ای سازوار بر اساس آماره های تربیی، نمونه های اولیه را با روش نمونه گیری تصادفی ساده بدون جایگذاری و با اندازه های $n = 5, 10, 15, 20$ از جمعیت شبیه سازی شده، انتخاب می کنیم.

برای هر n ، مقدار $\frac{2}{n}$ از ۱ تا $\frac{2}{n}$ تغییر می کند. برای هر ترکیب n ، ۱۰۰۰ مرتبه نمونه گیری را تکرار می کنیم و برای هر تکرار

با افزایش مقدار $RE(HT)$ ، افزایش می یابد (در برخی از جامعه ها مانند $N=10 \times 10$ و $N=20 \times 10$ با $n=5$). با کاهش مقدار c ، $E(v')$ افزایش یافته و بنابراین واریانس برآورده میانگین در نمونه گیری تصادفی ساده کاهش می یابد. در هر یک از جامعه ها، با افزایش اندازه نمونه اولیه (n)، مقادیر MSE برای هر دو برآورده HT و HH کاهش یافته است و علاوه بر آن معمولاً افزایشی در مقادیر RE ، برای هر دو برآورده HT و HH مشاهده می شود. مقادیر MSE برآورده های سازوار با افزایش n افزایش یابد، این کاهش چشمگیرتر است. انتخاب کوچک واحدهای اولیه نمونه گیری، MSE برآورده های سازوار را کاهش داده است (مقایسه مقادیر MSE برای دو جامعه $N=10 \times 10$ و $N=10 \times 20$).

بنابراین همان طور که انتظار داشتیم، در جامعه های خوش ای تر، برآورده های سازوار برای میانگین جامعه، بهتر از برآورده HT میانگین نمونه در نمونه گیری تصادفی ساده عمل می کنند.

در این راستا مشکل و خطیر شود به طوری که هر آماردانی همزمان باید کاربران علوم دیگر را در ک کرده و در عین حال به اصول آمار و احتمال پای بند بماند.

کاربرانی همچون زیست محیط شناسان با وجود این که استفاده از آزمونهای معنی دار آماری را به عنوان یک معیار مهم در تصمیم گیری هایشان در نظر می گیرند، اما معتقدند که بزرگترین خطایی که هر زیست محیط شناس می تواند در به کار بردن آمار مرتكب شود این است که مفهوم معنی دار بودن آماری را با معنی دار بودن زیست محیطی اشتباه بگیرد! این جمله در کتاب [۵] آمده است. خواننده علاقمند را به خواندن فصل اول این کتاب ارجاع می دهیم.

در این مقاله یک مثال عملی از یک طرح نمونه گیری ارانه شد و همان گونه که دیدیم کاربران از برآوردهای واریانس برای ارزیابی طرح نمونه گیری استفاده کردند. اگرچه ارزیابی یک طرح نمونه گیری بر اساس برآورد واریانس می تواند نشانگر خاصیت و رفتار طرح نمونه گیری باشد ولی هرگز نمی تواند به عنوان مدرک مستدلی برای ارزیابی قابل دفاع از طرح نمونه گیری به کار رود.

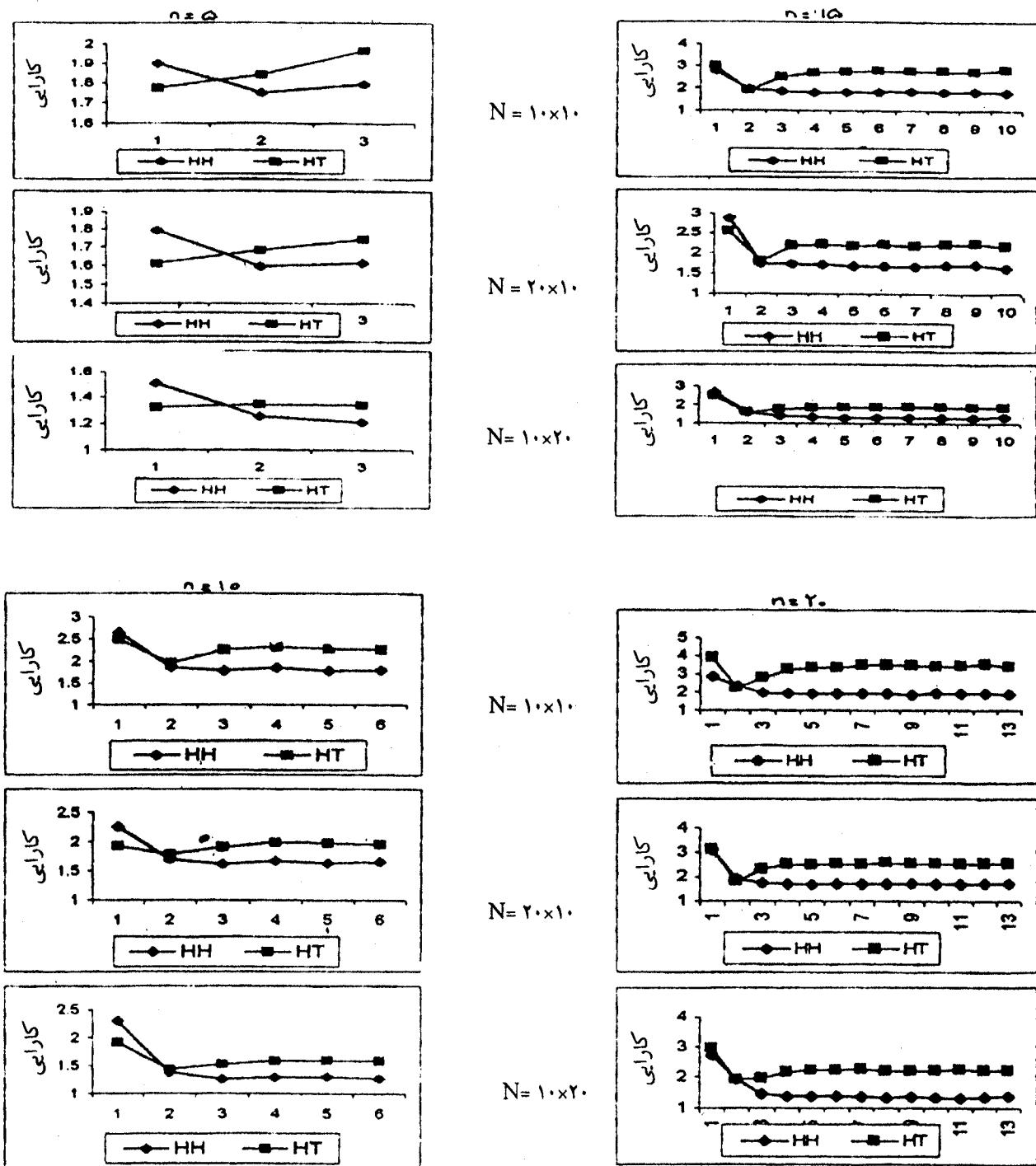
نویسنده گان این مقاله معتقدند که تا این مرحله که کاربران پیش رفته اند، بسیار امیدوار کننده است و اینک وظیفه آماردان است که با استفاده از نظریه و شبیه سازی به این سؤال پاسخ دهند که نتیجه کاربران که بر اساس یک نمونه به دست آمده است، چقدر صحت دارد. به همین دلیل در این مقاله بعد از عنوان یک مثال عملی، یک مثال شبیه سازی شده ارانه شد تا نتایج عملی به دست آمده، مورد بحث و بررسی قرار گیرد. نتایج این بررسی به طور خلاصه به شرح زیر است:

برآورده HT - تامسون (HT) معمولاً کارتر از برآورده هنسن - هوروپتس (HH) در نمونه گیری سازوار است. مثال شبیه سازی شده نیز نشان داد که در بسیاری از موارد برآورده HT بهتر از برآورده HH عمل می کند و در بعضی از جامعه ها با افزایش مقدار c کارایی آن بیشتر از HH می شود. همچنین

جدول ۱: مقایسه برآوردهای سازوار HH و HT و برآوردگر فراوانی (مایل دریایی / کیلوگرم) در نمونه گیری تصادفی ساده

طبقه	آماره	Srs	HH	HT
ناحیه مورد مطالعه ماهیهای لوتوی اقیانوس آرام				
Stop	برآورد فراوانی	۴۲۲	۴۲۰	۴۲۱
	SE	۱۶۶	۱۴۹	۱۴۹
ناحیه مطالعه ماهیهای Rougheye و Shortraker				
Intersection	برآورد فراوانی	۱۴۶۱	۱۱۱۱	۴۶۵
	SE	۸۲۲	۵۱۱	۳۲۱
West Gully	برآورد فراوانی	۲۹۷	۲۹۱	۲۹۱
	SE	۱۳۳	۱۲۷	۱۲۷
East Gully	برآورد فراوانی	۳۰۴	۳۴۲	۳۴۳
	SE	۲۱۳	۲۰۲	۲۰۲
ناحیه مطالعه ماهیهای North				
	برآورد فراوانی	۱۳۷۶	۱۸۸۴	۱۸۸۷
	SE	۲۹۳	۵۹۳	۵۹۲
South	برآورد فراوانی	۰۱۶	۰۱۷	۰۱۷
	SE	۱۲۸	۱۲۹	۱۲۹

شکل ۴: کارایی برآوردگرهای سازوار نسبت به برآوردگر میانگین در نمونه گیری تصادفی ساده در سه جامعه با اندازه های نمونه اولیه $n = 5, 10, 15, 20$



منابع

- [1] Thompson, S. K., (1990), *Adaptive Cluster Sampling*, J.A.M., Statist. Assoc., 85, 1103-1115.
- [2] Thompson, S. K., (1996), *Adaptive Cluster Sampling Based on Order Statistics*, Environmetrics, Vol. 7, 123-133.
- [3] Heifetz, J., Ianelli, J. N. and Clausen, D. M., (1996), *Slope Rockfish*, In: Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska as Projected For 1997.
- [4] Quinn II, T. J., Hanselman, D. H., Clausen, D. M., Heifetz, J. and Lunsford, C., (1999), *Adaptive Cluster Sampling of Rockfish Population*, Procceding of American Statistical Association, Joint Statistical Meeting, Biometrics Section 11-20.
- [5] Krebs, Chales J., (1998), *Ecological Methodology*.

آمار مطالعه لذت بخشی است در باب این موضوع که چگونه می توان جهان ناشناخته را با گشودن چند دریچه بر روی آن توصیف کرد.