نمونه گیری خوشه ای سازوار بر اساس آماره های ترتیبی

محمد صالحی [۱]　　فرخنده السادات سجادی [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 نمونهٔ 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 کنند، کارایی این روش زیاد می شود. همچنین تأثیر اندازهٔ واحدهای اولیهٔ نمونه گیری در میزان کارایی نیز، بررسی می شود.

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

۱. مقدمه

معمولاً بیشتر توجه در نظریهٔ 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 نمونهٔ کامل واحدها را قبل از عملیات میدانی مشخص کرد. اما در بسیاری از وضعیتهای نمونه گیری، مایلیم تصمیمهای خود را در طول مطالعه، بر اساس آنچه تاکنون مشاهده کرده ایم، اتخاذ کنیم.

نمونه گیری سازوار، نمونه گیریی است که در آن انتخاب واحدهایی که در نمونه قرار می گیرند، بستگی به مقادیر متغیر مورد علاقهٔ مشاهده شده در طول مطالعه دارد. یکی از اهداف نمونه گیری سازوار این است که از مشخصه های مفید جامعهٔ مورد مطالعه استفاده کرده تا برآوردی دقیقتر برای پارامترهای جامعه، مانند چگالی و یا فراوانی جامعهٔ، 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 وسیلهٔ 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

---

۱ دانشکدهٔ علوم ریاضی، دانشگاه صنعتی اصفهان

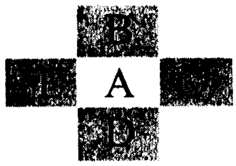
۲ دانشجوی کارشناسی ارشد دانشکدهٔ علوم ریاضی، دانشگاه صنعتی اصفهان

۳ Thompson

[Downloaded from andisheyeamari.irstat.ir on 2025-05-21]

(بدون جایگذاری ـ با جایگذاری)، نمونه گیری سیستماتیک و ... یک نمونهٔ اولیه از واحدهای جامعه انتخاب می شود، و هر گاه متغیر مورد مطالعهٔ 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 رابطهٔ همسایگی، یک رابطهٔ متقارن است. بدین معنا که اگر واحد i در همسایگی واحد j باشد، آنگاه واحد j نیز در همسایگی واحد i است. این همسایگیها به مقادیر متغیر مورد مطالعهٔ جامعه بستگی ندارند. همچنین این همسایگیها ممکن است بر اساس یک رابطهٔ خاص که بین واحدها وجود دارد، تعریف شوند. همسایگی می تواند به صورت یک منطقهٔ جغرافیایی نیز در نظر گرفته شود. در شکل ۱، همسایه های واحد A ، واحدهای B ، C ، D و E 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 همسایهٔ واحد نمونهٔ اولیه)، به وسیلهٔ یک فاصله یا یک مجموعهٔ C در برد متغیر مورد مطالعه معین می شود.

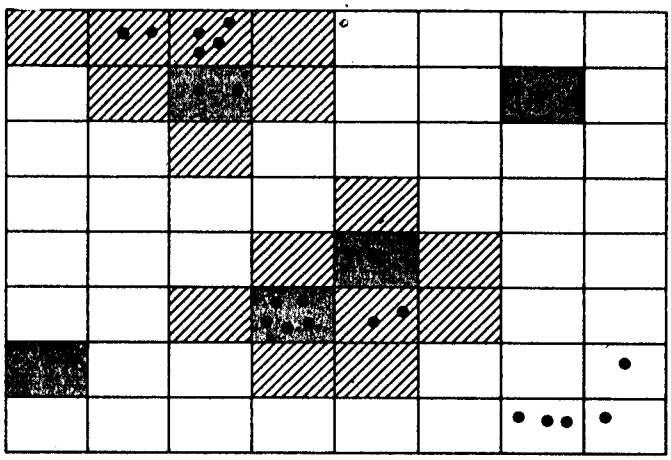
فرض می کنیم از یک جامعهٔ متناهی با اندازهٔ N ، نمونه گیری می کنیم و اعضا را با برچسبهای ۱،۲،...،N مشخص کرده و مقادیر مورد علاقه را به صورت $y = \{y_1, y_2, ..., y_N\}$ نشان می دهیم، به طوری که y_i مقدار متغیر مورد مطالعه برای واحد i ام است. در این صورت می گوییم واحد i در شرط مورد نظر که آن را با C نمایش 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، همهٔ واحدهایی که در همسایگی آن هستند 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 آید. 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 نمونهٔ اولیه انتخاب شده باشد، تشکیل یک شبکه با اندازهٔ ۱ را می دهد. در این طرح نمونه گیری ممکن است تعدادی قابل توجه از واحدهای نمونهٔ اولیه، در شرط از پیش مشخص شدهٔ 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 دریایی 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 ناحیهٔ مورد مطالعه با ٦٤ واحد را نشان می دهد. یک نمونهٔ تصادفی اولیه با ٤ واحد (واحدهای خاکستری) در شکل نشان داده شده است. واحدهای همسایه به نمونه اضافه می شوند، هر گاه حداقل یک شیء از جامعه در واحد انتخاب شده مشاهده شود. نمونهٔ نهایی شامل واحدهای خاکستری و واحدهای هاشور زده شده است.

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

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

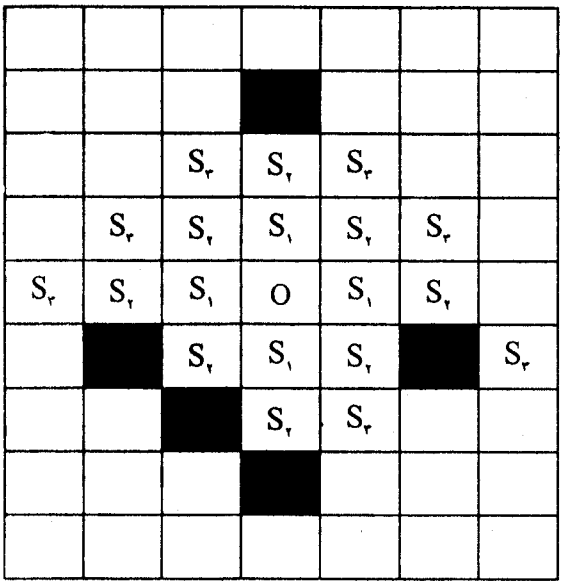
[Downloaded from andisheyeamari.irstat.ir on 2025-05-21]



شکل ۲

نمونه، به صورت $y_i > c$ است که c مقدار ثابتی است. اگر c مقدار نامناسب کوچکی انتخاب شود، ممکن است تعداد واحدهای لازم برای اضافه شدن به مجموعهٔ 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 عملاً نمونهٔ نهایی شامل همان نمونهٔ اولیه باشد. در این گونه موارد تامسون در سال ۱۹۹۶ طرح نمونه گیری سازوار خوشه ای بر اساس آماره های ترتیبی را مطرح کرد [۲]. بدین ترتیب که فرض کنید در نمونه گیری اولیه با اندازهٔ n، مقادیر $y_1, y_2, \ldots, y_n$ را برای واحدهای $1, 2, \ldots, n$ مشاهده کنیم. آنها را به صورت $y_{(1)} \le y_{(2)} \le \ldots \le y_{(n-r)} \le \ldots \le y_{(n)}$ مرتب می کنیم به طوری که $y_{(i)}$، مقدار i امین آمارهٔ ترتیبی باشد. برای هر واحد i در نمونه، همسایه های آن، به نمونه اضافه می شوند، اگر y_i ی مربوط به آن واحد از مقدار $(n - r)$ امین آمارهٔ ترتیبی بزرگتر باشد.

استفاده از نمونه گیری سازوار بر اساس آماره های ترتیبی، کنترل قابل توجهی روی حجم نمونهٔ نهایی دارد. اما اگر با یک جامعهٔ بسیار غیر طبیعی با خوشه های بزرگ روبه رو شویم، بعد از انتخاب نمونهٔ اولیه، باز هم مجبور به انتخاب واحد های زیادی هستیم که ممکن است برای ما عملی نباشد. به همین دلیل بعضی از استفاده کنندگان این طرح، از قبل، تعداد مراحلی را که به نمونهٔ سازوار اضافه می شود مشخص می کنند و بعد از آن نمونه گیری را متوقف می کنند. قاعدهٔ توقف S را به عنوان سطح S ام نمونه گیری سازوار ـ جدا از واحد انتخاب شده در نمونه گیری اولیه ـ تعریف می کنیم که بعد از این سطح، فرایند نمونه گیری متوقف می شود.



شکل ۳

شکل ۳ نمونه گیری خوشه ای سازوار با قاعدهٔ توقف $S = ۳$ است. (O نشان دهندهٔ واحد نمونهٔ اولیه و واحدهای سایه زده شده، نشان دهندهٔ واحدهای مرزی اند.)

٤. برآوردگرهای هَنسن ـ هورویتس[٤] و هورویتس ـ تامسون[٥]

یک برآوردگر نااریب برای میانگین، برآوردگر هَنسن ـ هورویتس است که برای نمونه گیری خوشه ای سازوار بر اساس آماره های ترتیبی به صورت زیر است:

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

---

[٤] Hansen - Hurwitz

[٥] Horvitz - Thompson

[Downloaded from andisheyeamari.irstat.ir on 2025-05-21]

که w_i و y_i^* به ترتیب میانگین و مجموع کل مشاهده در شبکه ای است که با واحد i ام نمونهٔ اولیه اشتراک دارد. برآوردگر هورویتس ـ تامسون برای میانگین، به صورت زیر است:

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

که y_k^* مجموع مقادیر y برای شبکهٔ k ام، κ تعداد شبکه های مجزا در نمونه، و α_k احتمال این است که شبکهٔ k در نمونه باشد. اگر x_k واحد، در شبکهٔ k ام وجود داشته باشند، آنگاه

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

در نمونه گیری خوشه ای سازوار بر اساس آماره های ترتیبی، برخلاف نمونه گیری خوشه ای سازوار، برآوردگر هورویتس ـ تامسون اریب است.

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

بررسیهای ماهیگیری (طی کردن مسیر مورد نظر در دریا با کشتی و شمارش ماهیها با صید کردن آنها و یا استفاده از روشهای مختلف دیگر) با طرح نمونه گیری سادهٔ 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 خوشه ای است (در مکانهایی خاص تجمع کرده اند)، ممکن است طرحی کارا نباشد. ماهیهای صخره ای[۶] در شمال شرقی اقیانوس آرام، نمونه ای از این گونه ها هستند. مجموعهٔ ماهیهای صخره ای در خلیج آلاسکا از گونه های مختلف تشکیل می شود.

[۶] Rockfish

ماهی لوتی اقیانوس آرام (POP)[۷] یکی از مهمترین گونه های تجارتی است. صید این گونه ماهی در سال ۱۹٦۵ به بالاترین مقدار خود (حدود ۳۵۰۰۰۰ تن) در دههٔ ۱۹٦۰ رسید. صید زیاد موجب کاهش نژاد این گونه ماهی شد. بررسیهای ماهیگیری در سرتاسر خلیج، هرسه سال یک بار، از سال ۱۹۸٤، برای برآورد بیومس[۸] ماهیهای صخره ای انجام شده است [۳]. در سال ۱۹۹۸ یک پروژهٔ مشترک برای یافتن برآوردهای بهتر از فراوانی ماهیهای صخره ای بین دانشگاه فربانکس آلاسکا[۹]، صنعت شیلات و آزمایشگاه ملی آک بی[۱۰]، در ناحیهٔ کوچکی از خلیج آلاسکا آغاز شد [٤]. فرضیهٔ 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. فرضیهٔ بعدی آنها عبارت بود از این که برآورد فراوانی POP از طریق نمونه گیری خوشه ای سازوار، سودمندتر از برآورد گونه های Shortraker و Rougheye از ماهیهای صخره ای است. به عبارت دیگر استفاده از نمونه گیری خوشه ای سازوار برای جمعیتهایی مانند Shortraker و Rougheye که خوشه ای نیستند، مفید نخواهد بود، زیرا آنهامعتقد بودند که POP ها در توزیعشان نسبت به Shortraker و Rougheye خوشه ای تر هستند. یک مطالعهٔ دو هفته ای در اوت ۱۹۹۸ بر طبق طرح نمونه گیری خوشه ای سازوار در ٦ 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

[۷] Pacific Ocean Perch

[۸] BIOMASS

[۹] University of Alaska Fairbanks

[۱۰] Auke Bay

[۱۱] Portlock Bank

[Downloaded from andisheyeamari.irstat.ir on 2025-05-21]

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

$$y_{(1)} \le y_{(2)} \le \ldots \le y_{(n-r)} \le \ldots \le y_{(n)}$$

که n، 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 $y_{(n-r)}$ باشد. برای ناحیهٔ مورد مطالعهٔ Shortraker-Rougheye، مقدار CPUE برای دو گونهٔ Shortraker و Rougheye ترکیب شد. در ابتدا مقدار r برابر با ۳ انتخاب شد.

در طبقات West Gully، East Gully و slop_Gully Intersection همسایگی به صورت چهار واحد در شمال - جنوب - شرق - غرب واحد انتخاب شده در نمونهٔ اولیه و در طبقات Slop و Shortraker_Rougheye، همسایگی به صورت خطی تعریف شد.

برای محدود کردن نمونه گیری سازوار، یک قاعدهٔ 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 حداکثر 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، حداکثر ۱۲ واحد سازوار مورد بررسی قرار گرفتند.

جدول ۱ برآوردهای به دست آمده برای فراوانی ماهیهای صخره ای و انحراف استاندارد (SE) این برآوردگرها را نشان می دهد. Srs نشان دهندهٔ برآورد به دست آمده از نمونه گیری تصادفی ساده است. HH و HT به ترتیب بیانگر مقادیر برآوردگرهای سازوار هَنسِن - هورویتس و هورویتس - تامسون هستند. در این جدول مشاهده می شود، نمونه گیری سازوار برای ماهیهای لوتی اقیانوس آرام بهتر از نمونه گیری تصادفی ساده در طبقات مختلف عمل کرده است. در مقایسه با نمونه گیری تصادفی ساده، مقدار SE برای برآوردگرهای سازوار از فراوانی ماهیهای لوتی اقیانوس آرام در هر ۴ طبقه کاهش یافته، اما نمونه گیری سازوار در مقایسه با نمونه گیری تصادفی ساده برای ماهیهای Shortraker و Rougheye در طبقهٔ شمال که توزیع جمعیت آنها به صورت خوشه ای نبوده، خوب عمل نکرده است. برای طبقهٔ 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 Shortraker و Rougheye 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 پایهٔ 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. مقایسهٔ کارایی نمونه گیری تصادفی سادهٔ و نمونه گیری خوشه ای سازوار بر اساس آماره های ترتیبی

فرض کنید ناحیهٔ مورد مطالعه را به ۴۰۰ واحد مربع تقسیم کرده ایم که گونهٔ 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:

---

[۱۲] Cattch-Per-Unit-Effort

[Downloaded from andisheyeamari.irstat.ir on 2025-05-21]

تعدادی گیاه مولد از توزیع پواسون با میانگین λ_p تولید کرده و با استفاده از توزیع یکنواخت در فاصلهٔ (۰,۲۰) مکان این گیاهان را در ناحیهٔ مورد نظر مشخص می کنیم. گیاهان فرعی هر گیاه مولد را بر طبق یک فرایند پواسون با میانگین λ_c تولید کرده و با استفاده از توزیع نرمال دومتغیره ای که میانگین آن مختصات مکان گیاه مولد در ناحیه و واریانس آن $\sigma^2 I_{2\times 2}$ است، در اطراف گیاه مولد پراکنده می کنیم. با مقادیر کوچک σ^2 خوشه هایی از گیاه را به دست می آوریم که ناحیهٔ کوچکی را می پوشانند. مقادیر بزرگ σ^2 خوشه هایی با پراکندگی زیاد در ناحیهٔ 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 مطالعهٔ چگونگی در نظر گرفتن اندازهٔ واحدهای اولیه، ناحیه را به سه حالت زیر مورد بررسی قرار می دهیم:

۱) ناحیه، مربعی به طول ضلع ۱۰ واحد مربع است ($N = ۱۰\times ۱۰$).

۲) ناحیه، مستطیلی به طول ۲۰ و عرض ۱۰ واحد مربع است ($N = ۲۰\times ۱۰$).

۳) ناحیه، مستطیلی به طول ۱۰ و عرض ۲۰ واحد مربع است ($N = ۱۰\times ۲۰$).

میانگین و واریانس سه جمعیت تولید شده عبارت اند از:

$$N = ۱۰\times ۱۰ : \mu = ۵/۸۵, \sigma^2 = ۷۶۲/۶۳۳۸$$
$$N = ۲۰\times ۱۰ : \mu = ۲/۹۲۵, \sigma^2 = ۲۵۸/۴۶۱۷$$
$$N = ۱۰\times ۲۰ : \mu = ۲/۹۲۵, \sigma^2 = ۲۰۴/۶۹۲۸$$

برای مقایسهٔ کارایی نمونه گیری تصادفی ساده و نمونه گیری خوشه ای سازوار بر اساس آماره های ترتیبی، نمونه های اولیه را با روش نمونه گیری تصادفی سادهٔ بدون جایگذاری و با اندازه های $n = ۵,۱۰,۱۵,۲۰$ از جمعیت شبیه سازی شده، انتخاب می کنیم. برای هر n، مقدار r از ۱ تا $\frac{۲}{۳}n$ تغییر می کند. برای هر ترکیب n و r، ۱۰۰۰۰ مرتبه نمونه گیری را تکرار می کنیم و برای هر تکرار مقادیر $\hat{\mu}_{HH}$ ، $\hat{\mu}_{HT}$ و اندازهٔ نمونهٔ 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 رابطهٔ زیر به دست می آوریم:

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

سرانجام مقادیر ($R = ۱۰۰۰۰,\ \hat{\mu}_i = \hat{\mu}_{HH}$ or $\hat{\mu}_{HT}$)

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

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

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

را برای هر برآوردگر سازوار هَنسِن ـ هورویتس یا هورویتس ـ تامسون به دست می آوریم. نمودار های شکل ٤ کارایی برآوردگرهای سازوار را نسبت به برآوردگر میانگین در نمونه گیری تصادفی ساده (بدون جایگذاری) نشان می دهند.

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 خوشه ای سازوار جمعیت ماهیهای صخره ای» گرفته شده است [٤]. ما سعی کرده ایم تا آنجا که ممکن است، مقالهٔ حاضر را از فرمول و نظریهٔ موضوع مورد مطالعه دور نگه داریم، به این امید که مجموعهٔ بزرگتری از خوانندگان این مجله را به خواندن مقاله ترغیب کنیم. به طور سنتی، جامعهٔ دانشگاهی آماری ایران، توجه خاصی به مبانی نظری و ریاضیات داشته است و این امر مایهٔ مباهات 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 وظیفهٔ آماردانان

[Downloaded from andisheyeamari.irstat.ir on 2025-05-21]

در این راستا مشکل و خطیر شود به طوری که هر آماردانی همزمان باید کاربران علوم دیگر را درک 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 مرتکب 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. خوانندهٔ علاقمند را به خواندن فصل اول این کتاب ارجاع می دهیم.

در این مقاله یک مثال عملی از یک طرح نمونه گیری ارائه 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.

نویسندگان این مقاله معتقدند که تا این مرحله که کاربران پیش رفته اند، بسیار امیدوار کننده است و اینک وظیفهٔ آماردان است که با استفاده از نظریهٔ و شبیه سازی به این سؤال پاسخ دهند که نتیجهٔ 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 ارائه شد تا نتایج عملی به دست آمده، مورد بحث و بررسی قرار گیرد. نتایج این بررسی به طور خلاصه به شرح زیر است:

برآوردگر هورویتس ـ تامسون (HT) معمولاً کاراتر از برآوردگر هَنسِن ـ هورویتس (HH) در نمونه گیری سازوار است. مثال شبیه سازی شده نیز نشان داد که در بسیاری از موارد برآوردگر HT بهتر از برآوردگر HH عمل می کند و در بعضی از جامعه ها با افزایش مقدار r کارایی آن بیشتر از HH می شود. همچنین با افزایش مقدار r، RE(HT) افزایش می یابد (در برخی از جامعه ها مانند N =۱۰×۱۰ و N =۲۰×۱۰ با n =۵). با کاهش مقدار c، $E(v')$ افزایش یافته و بنابراین واریانس برآوردگر میانگین در نمونه گیری تصادفی ساده کاهش می یابد. در هر یک از جامعه ها، با افزایش اندازهٔ نمونهٔ اولیه (n)، مقادیر MSE برای هر دو برآوردگر HT و HH کاهش یافته است و علاوه بر آن معمولاً افزایشی در مقادیر RE، برای هر دو برآوردگر سازوار HT و HH مشاهده می شود. مقادیر MSE برآوردگرهای سازوار با افزایش حالت خوشه ای جامعه ها، کاهش یافته و هرچه اندازهٔ نمونهٔ اولیه (n) افزایش یابد، این کاهش چشمگیرتر است. انتخاب کوچک واحدهای اولیهٔ نمونه گیری، MSE برآوردگرهای سازوار را کاهش داده است (مقایسهٔ مقادیر MSE برای دو جامعهٔ N =۱۰×۱۰ و N =۱۰×۲۰).

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

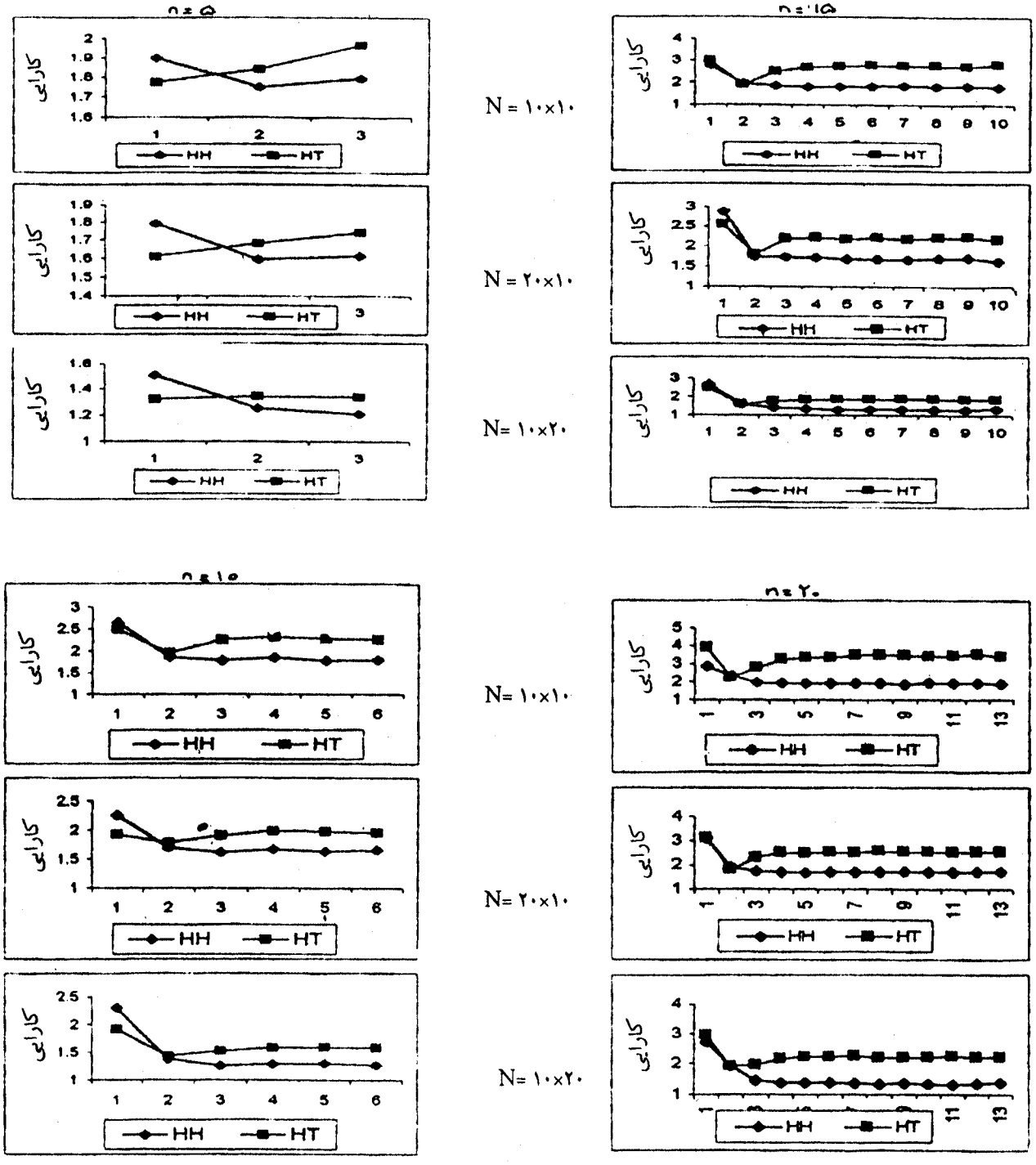
[Downloaded from andisheyeamari.irstat.ir on 2025-05-21]

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

طبقه	آماره	Srs	HH	HT
ناحیهٔ مورد مطالعهٔ ماهیهای لوتی اقیانوس آرام				
Stop				
	برآورد فراوانی	۴۲۲	۴۲۰	۴۲۱
	SE	۱۶۶	۱۴۹	۱۴۹
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	۳۵۴	۳۴۲	۳۴۳
	SE	۲۱۳	۲۰۲	۲۰۲
ناحیهٔ مورد مطالعهٔ ماهیهای Shortraker و Rougheye				
North				
	برآورد فراوانی	۱۳۷۶	۱۸۸۴	۱۸۸۷
	SE	۲۹۳	۵۹۳	۵۹۲
South				
	برآورد فراوانی	۵۱۶	۵۱۷	۵۱۷
	SE	۱۲۸	۱۲۹	۱۲۹

[Downloaded from andisheyeamari.irstat.ir on 2025-05-21]

شکل ٤: کـــارایی برآوردگرهای سازوار نسبت به برآوردگر میانگین در نمونه گیری تصادفی ساده در سه جامعه با اندازه های نمونهٔ اولیهٔ $n = ۵,۱۰,۱۵,۲۰$



[Downloaded from andisheyeamari.irstat.ir on 2025-05-21]

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 Statitical Meeting, Biometrics Section 11-20.

[5] Krebs, Chales J., (1998), *Ecological Methodology.*

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

[Downloaded from andisheyeamari.irstat.ir on 2025-05-21]