مجله گزیده‌مطالب آماری
سال ۱۸، شماره‌ی ۲، پاییز و زمستان ۱۳۸۶، صص ۹۵-۸۳

کاربرد روش برآورد بیز تجربی برای ناحیه‌ی کوچک در نمونه‌گیری چرخشی

علیرضا خوشگویان‌فرد،[۱،*] حمیدرضا نواب‌پور،[۲] محمدرضا نمازی راد[۳]

[۱] مرکز تحقیقات صدا و سیما
[۲] دانشگاه علامه طباطبایی
[۳] دانشگاه ولنگانگ، استرالیا

چکیده. آمارگیری مکرر روشی برای مطالعه‌ی الگوی تغییرات یک پارامتر در طول زمان است. آمارگیری چرخشی نوعی از آمارگیری مکرر است که نه‌تنها برای بررسی روندها، بلکه برای یافتن برآورد دقیق‌تر یک پارامتر در یک بازه‌ی زمانی هنگامی که در طول این بازه دستخوش تغییرات می‌شود، به کار می‌رود. استفاده از آمارگیری چرخشی نیازمند صرف هزینه و زمان قابل توجهی است به‌ویژه هنگامی که دوره‌های زمانی دارای فاصله‌ی کوتاهی از یکدیگر باشند، برای مثال آمارگیری‌های ماهانه. از این رو، استفاده از راهکارهایی که به کاهش اندازه‌ی نمونه منجر شوند در آمارگیری چرخشی راهگشا خواهند بود. در این مقاله، همزمان با بهره‌گیری از یک روش برآورد برای ناحیه‌ی کوچک در آمارگیری چرخشی، تلاش می‌شود راهکاری برای کاهش اندازه‌ی نمونه ارائه شود. با استفاده از مطالعه‌ی شبیه‌سازی، این راهکار با شیوه‌ی کلاسیک (بدون کاهش نمونه) مقایسه شده است.

۱- مقدمه

افزایش جمعیت، تعدد نیازها و گوناگونی حوزه‌های تصمیم‌گیری، دولت‌ها را به استفاده از شاخص‌های مختلف اقتصادی، اجتماعی و فرهنگی ترغیب کرده و مراکز آماری را به

واژگان کلیدی: آمارگیری مکرر؛ آمارگیری چرخشی؛ برآورد برای ناحیه‌ی کوچک؛ شبیه‌سازی به‌روش مونت‌کارلو؛ کارایی.
* نویسنده‌ی عهده‌دار مکاتبات

اجـرای آمـارگیری‌هـای مختلـف بـرای تولیـد آمارهـای مـورد نیـاز سـوق داده اسـت. آمارگیری‌های نمونه‌ای[۱] از جمله متداول‌ترین فعالیت‌های آماری هستند که اطلاعات مورد نیاز را برای تصمیم‌گیری در سطوح خرد و کلان فراهم مـی‌کننـد. در ایـن میـان، اجـرای برخی از آمارگیری‌ها موسوم به آمارگیری‌های مکرر،[۲] در فاصله‌های زمـانی مـنظم تکـرار می‌شوند.

معمولاً آمارگیری‌های مکرر یا به‌منظور دست‌یابی به روند تغییرات یک شاخص در طول زمان اجرا می‌شوند یا برای برآورد دقیق‌تر پارامتری در یک بازه‌ی زمانی مشخص (مثلاً یک سال) که در طول این بازه دستخوش تغییرات می‌شود. سه روش عمومی برای اجرای آمارگیری به‌صورت مکرر وجود دارد: آمارگیری مقطعی،[۳] آمارگیری پانلی[۴] و آمارگیری چرخشی.[۵] در روش نخست، نمونه‌ی هر دوره‌ی زمانی، مستقل از نمونه‌ی دوره‌ی دیگر است در حالی که در روش دوم، به یک نمونه‌ی ثابت در دوره‌های زمانی مختلف مراجعه می‌شود. آمارگیری چرخشی، تلفیقی از دو روش پیش‌گفته است یعنی برخی از واحدهای نمونه‌ای دوره‌های آمارگیری، یکسان و برخی متفاوت هستند. خواننده می‌تواند بحث تفصیلی در باره‌ی این سه روش و مقایسه‌ی کارایی آن‌ها را در [۲] بیابد.

اگرچه آمارگیری مکرر می‌تواند به آشکار شدن روند تغییرات یک شاخص در طول زمان کمک کند ولی اجرای پیوسته‌ی آن در دوره‌های زمانی مختلف (مثلاً فصلی) و به‌منظور برآورد آماره‌هایی در سطوح جغرافیایی غیر ملی (مثلاً استانی) امری هزینه‌بر و زمان‌بر است زیرا به افزایش اندازه‌ی نمونه می‌انجامد. بنا بر این، به‌کارگیری آن در کنار روش‌های برآورد برای ناحیه‌ی کوچک[۶] می‌تواند شرایط بهینه‌ای را فراهم کند. خواننده می‌تواند در باره‌ی مسئله برآورد برای ناحیه‌ی کوچک و قابلیت‌های آن به [۴] مراجعه کند.

در این مقاله به‌کمک شبیه‌سازی، کارایی یک روش برآورد برای ناحیه‌ی کوچک برای برآورد تفاضل دو نسبت در حالی که داده‌ها از آمارگیری چرخشی تولید شده‌اند، ارزیابی می‌شود. این ارزیابی تحت تداخل‌های مختلف اجرا می‌شود تا نقش آن نیز در کارایی این برآوردگر بررسی شود. یافته‌ها در دو بخش ارائه می‌شوند. ابتدا طرح مطالعه‌ی شبیه‌سازی شامل معرفی اجمالی برآوردگر، ساختار جامعه‌ی آماری و فرایند شبیه‌سازی شرح داده می‌شود و پس از آن، یافته‌ها و بحث در باره‌ی آن‌ها ارائه خواهد شد.

۲- مطالعه‌ی شبیه‌سازی

در این بخش، ابتدا برآوردگر بیز تجربی برای برآورد پارامتری از نوع نسبت (درصد) به‌عنوان یک روش برآورد برای ناحیه‌ی کوچک به‌اختصار معرفی می‌شود. سپس ساختار و چگونگی تولید جامعه‌ی آماری فرضی که مبنای شبیه‌سازی داده‌ها است، معرفی می‌شود. سرانجام، فرایند اجرای شبیه‌سازی شامل طرح نمونه‌گیری، اندازه‌ی نمونه و معیار مقایسه‌ی برآوردها شرح داده می‌شود.

۲-۱- معرفی برآوردگر ناحیه‌ی کوچک

برآوردگرهای مدل مبنا[۷] از اهمیت ویژه‌ای در برآورد برای ناحیه‌های کوچک برخوردارند زیرا ظرفیت استفاده از متغیرهای کمکی را بیش از برآوردگرهای طرح مبنا[۸] دارا هستند. شایان ذکر است که متغیرهای کمکی نقش عمده‌ای در جبران کمبود داده‌های نمونه‌ای در سطح ناحیه‌های کوچک ایفا می‌کنند. به عبارت دیگر، از طریق یک مدل می‌توان پارامتر مورد نظر را با دقت نسبتاً قابل قبولی برای یک ناحیه برآورد کرد در حالی که تعداد نمونه‌ی کافی از آن ناحیه در دست نیست تا مستقیماً بر اساس آن‌ها برآوردی ارائه داد.

انواع مختلفی از مدل‌ها برای برآورد برای ناحیه‌ی کوچک مورد استفاده قرار می‌گیرند. غالباً مدل‌سازی برای نسبت‌ها به‌کمک مدل‌های (خطی یا ناخطی) تعمیم‌یافته صورت می‌گیرد. در این‌جا از برآوردگر بیز تجربی استفاده می‌کنیم که مبتنی بر مدل آمیخته‌ی[۹] زیر است:

$$\rho = X\beta + u + e$$
$$u \sim N(\circ\,,\ \sigma_u^{۲}\mathbf{I}_K) \qquad e \sim N(\circ\,,\ diag(\sigma_۱^{۲},\ldots,\sigma_K^{۲})) \qquad (۱)$$

که در آن $diag$ نمایانگر یک ماتریس قطری، $\rho' = (\ln\frac{\hat{p}_۱^D}{۱-\hat{p}_۱^D},\ldots,\ln\frac{\hat{p}_K^D}{۱-\hat{p}_K^D})$ برداری از لوجیت‌های نسبت ناحیه‌های کوچک یعنی $\hat{p}_i^D$ ها، X ماتریس متغیرهای کمکی و β برداری از پارامترهای نامعلوم است. بر این اساس، برآوردگر بیز تجربی نسبت ناحیه‌ی کوچک iام به‌شکل زیر است:

$$\hat{P}_i^{EB} = \frac{\exp((1-\gamma_i)\rho_i + \gamma_i \bar{X}_i'\hat{\beta})}{1+\exp((1-\gamma_i)\rho_i + \gamma_i \bar{X}_i'\hat{\beta})} \quad (۲)$$

که در آن $\gamma_i = \sigma_i^2/\sigma_i^2 + \sigma_u^2$ و $\bar{X}_i'$ سطر iام ماتریس X و ρ_i مؤلفه iام بردار ρ است.

برآورد پارامترهای مدل آمیخته یعنی σ_i^2، σ_u^2 و بردار β بر اساس داده‌های موجود و به‌کمک روش‌های عددی تکراری انجام می‌شود. برای این منظور، ابتدا برآوردی از σ_i^2ها با روش دلتا[۱۰] به‌صورت زیر به دست می‌آید:

$$\hat{\sigma}_i^2 = \left(\frac{d\rho_i}{dp_i}\right)^2 \hat{V}(\hat{P}_i^D) \quad (۳)$$

در مرحله‌ی بعد، σ_u^2 برآورد می‌شود تا به این ترتیب برآوردی از ماتریس $V = diag(\sigma_1^2 + \sigma_u^2, \ldots, \sigma_K^2 + \sigma_u^2)$ و ضرایب مدل به دست آید. برای این منظور، با روش گشتاوری در قالب الگوریتم زیر یک برآورد اولیه برای σ_u^2 به دست می‌آوریم:

- ابتدا برآوردی از β مثلاً با روش کمترین توان‌های دوم معمولی OLS تعیین می‌شود یعنی $\hat{\beta}^{\text{OLS}} = (X'X)^{-1}X'\rho$.
- به‌کمک این برآورد، مانده‌های مدل یعنی $e = \rho - X\hat{\beta}$ محاسبه می‌شوند.
- قرار می‌دهیم $\nu = (\hat{\sigma}_1^2, \ldots, \hat{\sigma}_K^2)'$ و $s = \nu - X(X'X)^{-1}X\nu$ که در این صورت، برآورد گشتاوری عبارت است از $\hat{\sigma}_u^2 = \frac{e'e - s's}{K-L}$.

اکنون می‌توان با حل همزمان معادلات زیر به‌صورت تکراری به برآوردهای نهایی σ_u^2 و ضرایب مدل دست یافت:

$$\begin{cases} \hat{\sigma}_u^2 = \dfrac{(\rho - X\beta)'V^{-1}(\rho - X\beta)}{K-L} \\ \hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}\rho \end{cases} \quad (۳)$$

همان‌طور که دیده می‌شود پارامترهای σ_u^2 و σ_i^2ها با استفاده از اطلاعات داده‌های نمونه‌ای موجود محاسبه می‌شوند لذا (۲) را برآوردگر بیز « تجربی » می‌نامند. بحث بیش‌تر در باره‌ی این برآوردگر در [۱] ارائه شده است.

۲-۲- تولید داده‌ها

برای اجرای شبیه‌سازی، جامعه‌ای فرضی با یکصد هزار واحد و شامل ۱۶ زیرجامعه تولید شد. هر واحد از این جامعه دارای سه مقدار است: مقادیر Y_1 و Y_2 به‌عنوان مقادیر متغیر تحت مطالعه در دو دوره‌ی زمانی و مقدار X به‌عنوان مقدار متغیر کمکی مرتبط با متغیر Y. برای توجیه استفاده از آمارگیری چرخشی، مقادیر Y_1 و Y_2 با یکدیگر دارای همبستگی هستند. شایان ذکر است متغیر Y، دوحالتی (صفر-یک) و پارامتر مورد برآورد، تفاضل درصد در دو دوره‌ی زمانی یعنی $P_2 - P_1$ است.

جدول ۱ شامل مشخصات آماری مربوط به متغیر Y در کل جامعه و ۱۶ زیرجامعه‌ی آن است. همان‌طور که در سطر آخر جدول دیده می‌شود، کل جامعه شامل یکصد هزار واحد است که درصد « ۱ها » برای دوره‌ی اول برابر با ۱۱/۷۲ با واریانس ۱۰۳۴/۸۸۱ و برای دوره‌ی دوم برابر با ۱۹/۶۸ با واریانس ۱۵۸۰/۹۵۶ است. همچنین برای مثال، زیرجامعه‌ی ششم دارای ۲۵۰۰ واحد با درصد « ۱های » ۹/۲۴ و واریانس ۸۳۸/۹۵۸ در دوره‌ی اول و درصد ۱۴/۶۰ و واریانس ۱۲۴۷/۳۳۹ در دوره‌ی دوم است. گفتنی است زیرجامعه‌ها از توزیع تصادفی یکصد هزار واحد جامعه‌ی اصلی ایجاد شده‌اند (از آن‌جا که در جدول ۱، به‌جای نسبت از درصد برای ارائه‌ی ارقام استفاده شده است، مقادیر واریانس‌ها تا این حد بزرگ هستند).

جدول ۱- مشخصات آماری کل جامعه‌ی فرضی و زیرجامعه‌های آن

ناحیه	جمعیت	دوره‌ی اول		دوره‌ی دوم	
		درصد	واریانس	درصد	واریانس
۱	۲۰۰۰	۷/۰۵	۶۵۵/۶۲۵	۱۰/۸۵	۹۶۷/۷۶۱
۲	۳۰۰۰	۲۰/۹۳	۱۶۲۵/۶۸۱	۳۶/۱۷	۲۳۰۹/۴۰۹
۳	۲۵۰۰	۱۷/۸۰	۱۴۶۳/۷۴۵	۲۹/۲۴	۲۰۶۹/۸۵۰
۴	۴۰۰۰	۱۵/۹۸	۱۳۴۲/۶۳۵	۲۶/۶۵	۱۹۵۵/۲۶۶
۵	۶۰۰۰	۹/۰۸	۸۲۵/۹۶۴	۱۴/۷۵	۱۲۵۷/۶۴۷
۶	۲۵۰۰	۹/۲۴	۸۳۸/۹۵۸	۱۴/۶۰	۱۲۴۷/۳۳۹
۷	۲۰۰۰۰	۱۴/۳۷	۱۲۳۰/۸۲۱	۲۴/۳۹	۱۸۴۳/۹۶۴
۸	۱۰۰۰۰	۱۳/۴۳	۱۱۶۲/۷۵۱	۲۲/۴۲	۱۷۳۹/۵۱۸
۹	۵۰۰۰	۱۲/۳۶	۱۰۸۳/۴۴۷	۱۹/۹۲	۱۵۹۵/۵۱۳
۱۰	۷۵۰۰	۶/۰۳	۵۶۶/۴۲۱	۹/۹۷	۸۹۷/۹۸۶
۱۱	۲۵۰۰	۶/۹۲	۶۴۴/۳۷۱	۱۲/۳۲	۱۰۸۰/۶۵۰
۱۲	۲۰۰۰	۲۳/۰۰	۱۷۷۱/۸۸۶	۳۹/۳۰	۲۳۸۶/۷۰۳
۱۳	۳۰۰۰	۲۰/۷۳	۱۶۴۴/۰۱۰	۳۴/۶۰	۲۲۶۳/۵۹۵
۱۴	۱۰۰۰۰	۳/۴۲	۳۳۰/۳۳۷	۶/۰۳	۵۶۶/۶۹۶
۱۵	۸۰۰۰	۱۲/۲۵	۱۰۷۵/۰۷۲	۲۰/۷۸	۱۶۴۶/۱۰۵
۱۶	۱۲۰۰۰	۱۰/۲۴	۹۱۹/۳۵۲	۱۷/۲۹	۱۴۳۰/۲۸۴
کل	۱۰۰۰۰۰	۱۱/۷۲	۱۰۳۴/۸۸۱	۱۹/۶۸	۱۵۸۰/۹۵۶

۳-۲- فرایند شبیه‌سازی

برای اجرای شبیه‌سازی، ابتدا اندازه‌ی نمونه‌ای برابر با ۶۷۱ واحد بر اساس روش نمونه‌گیری تصادفی ساده بدون جایگذاری با حاشیه‌ی خطای ۵ درصد و سطح اطمینان ۹۵ درصد به‌منظور برآورد تفاضل درصدهای دو دوره‌ی زمانی در سطح « کل جامعه » محاسبه شد. بدیهی است با این تعداد نمونه قادر خواهیم بود برآورد قابل قبولی از $P_۲ - P_۱$ برای کل جامعه به دست آوریم. با وجود این، هدف اصلی، برآورد تفاضل درصدها برای تک‌تک زیرجامعه‌ها است در حالی که تنها همین ۶۷۱ نمونه در دست باشد.

بنا بر این، این تعداد نمونه متناسب با جمعیت زیرجامعه‌ها در بین آن‌ها توزیع شد تا از هر زیرجامعه تعدادی نمونه وجود داشته باشد.

جدول ۲ شامل دو اندازه‌ی نمونه برای هر زیرجامعه است. اول، تعداد نمونه‌ی تخصیص‌یافته به هر زیرجامعه با روش تخصیص متناسب. برای مثال، از ۶۷۱ نمونه، تعداد ۱۳ نمونه به زیرجامعه‌ی اول و تعداد ۵۰ نمونه به زیرجامعه‌ی دهم تخصیص یافته است. دوم، اندازه‌ی نمونه‌ی مورد نیاز برای هر زیرجامعه با حاشیه خطای ۵ درصد و سطح اطمینان ۹۵ درصد. این اندازه‌ی نمونه هنگامی اهمیت می‌یابد که با این پرسش روبه‌رو شویم که آیا می‌توان، برای مثال، به‌طور مستقیم بر اساس ۱۳ نمونه‌ی موجود از زیرجامعه‌ی اول، برآورد قابل قبولی از تفاضل درصدهای این زیرجامعه با حاشیه‌ی خطا و سطح اطمینان مورد نظر ارائه کرد؟ پاسخ منفی است زیرا برای برآورد مستقیم تفاضل درصدهای این زیرجامعه به‌طوری که حاشیه‌ی خطا و سطح اطمینان حفظ و واریانس آن نیز لحاظ شود به ۳۴۳ نمونه نیاز است در حالی که تنها ۱۳ نمونه از این زیرجامعه وجود دارد. بنا بر این، تعداد نمونه‌ی تخصیص‌یافته به زیرجامعه‌ی اول (۱۳ نمونه) بسیار کوچک‌تر از تعداد نمونه‌ی لازم (۳۴۳ نمونه) است. پس تحت این شرایط که اندازه‌ی نمونه برای کل جامعه و نه تک‌تک زیرجامعه‌ها بهینه شده است، هر یک از زیرجامعه‌ها به‌عنوان یک « ناحیه‌ی کوچک » قلمداد می‌شوند.

نکته‌ی قابل توجه در جدول ۲ به سطر آخر آن (کل) بر می‌گردد. در واقع، اگر برای هر یک از زیرجامعه‌ها، اندازه‌ی نمونه‌ی اختصاصی با توجه به حاشیه‌ی خطا، سطح اطمینان، واریانس و جمعیت آن‌ها تعیین شود، در مجموع به ۹۰۷۴ نمونه نیاز است. در حالی که اگر تنها یک اندازه‌ی نمونه در سطح کل جامعه محاسبه شود به ۶۷۱ نمونه نیاز خواهد بود. بنا بر این، اگر بتوان شیوه‌ای را به کار بست که بر اساس ۶۷۱ نمونه برآوردهای قابل قبولی برای زیرجامعه‌ها به دست دهد آن‌گاه قادر خواهیم بود در زمان و هزینه‌ی آمارگیری به‌طور چشمگیری صرفه‌جویی کنیم زیرا بدون انتخاب ۹۰۷۴ نمونه و تنها با ۶۷۱ نمونه به هدف خود خواهیم رسید.

جدول ۲- اندازه‌های نمونه در دو حالت

زیرجامعه	اندازه‌ی نمونه‌ی تخصیص‌یافته	اندازه‌ی نمونه‌ی لازم
۱	۱۳	۳۴۳
۲	۲۰	۷۴۲
۳	۱۷	۶۵۳
۴	۲۷	۶۹۱
۵	۴۰	۴۹۳
۶	۱۷	۴۳۹
۷	۱۳۴	۷۵۸
۸	۶۷	۷۳۰
۹	۳۴	۶۰۰
۱۰	۵۰	۳۶۵
۱۱	۱۷	۳۹۰
۱۲	۱۳	۶۷۵
۱۳	۲۰	۷۳۱
۱۴	۶۷	۲۳۷
۱۵	۵۴	۶۴۶
۱۶	۸۱	۵۸۱
کل	۶۷۱	۹۰۷۴

راهکاری که برای کاهش تعداد نمونه به کار می‌رود مبتنی بر دو ویژگی است: اول، استفاده از روش‌های برآورد برای ناحیه‌ی کوچک که حداکثر بهره‌برداری را از اطلاعات کمکی موجود برای جبران تعداد کم نمونه می‌کنند. دوم، استفاده از داده‌های حاصل از نمونه‌گیری چرخشی به‌جای دو نمونه‌ی مستقل است زیرا اشتراک در تعدادی از نمونه‌های دو دوره، خود می‌تواند به افزایش دقت برآورد تفاضل درصدها کمک کند. این مقاله به‌کمک شبیه‌سازی در صدد مقایسه‌ی این راهکار ترکیبی (برآوردگر ناحیه‌ی کوچک همراه با نمونه‌گیری چرخشی) با روش کلاسیک یعنی تنها به‌کارگیری نمونه‌گیری

چرخشی بر اساس محاسبه‌ی اندازه‌ی نمونه‌ی جداگانه برای هر زیرجامعه و برآورد مستقیم پارامترها بر اساس داده‌های هر زیرجامعه است.

٤-۲- معیار مقایسه‌ی برآوردها

همان‌طور که قبلاً اشاره شد، هدف، برآورد تفاضل درصدها برای یک متغیر دوحالتی در دو دوره‌ی زمانی است. این کار نیز با دو طرح متفاوت اجرا می‌شود. در حالت اول، اندازه‌ی نمونه برای کل جامعه تعیین و سپس با روش متناسب با جمعیت بین ۱۶ زیرجامعه توزیع می‌شود. پس از آن با استفاده از یک روش برآورد برای ناحیه‌ی کوچک (بیز تجربی) درصدهای هر ناحیه برآورد و تفاضل‌ها محاسبه می‌شوند. این فرایند، ۱۰۰۰ بار تکرار می‌شود لذا در پایان، برای هر ناحیه ۱۰۰۰ برآورد تفاضل وجود دارد.

از آن‌جا که داده‌ها به یک جامعه‌ی فرضی تعلق دارند، مقادیر واقعی تفاضل درصدها برای هر زیرجامعه معلوم است. به این ترتیب، می‌توان اریبی برآوردها را برای طرح ناحیه‌ی کوچک، از طریق رابطه‌ی زیر برای زیرجامعه‌ی j ام به دست آورد:

$$\text{Bias}_i = \frac{\sum_{j=1}^{1000}(\hat{\theta}_{ij} - \theta_i)}{1000} \qquad i = 1, 2, \ldots, 16 \tag{۴}$$

که در آن $\theta_i = P_{2i} - P_{1i}$ مقدار واقعی تفاضل درصدها برای زیرجامعه‌ی iام و $\hat{\theta}_{ij} = \hat{P}_{2ij} - \hat{P}_{1ij}$ برآورد تفاضل درصدهای زیرجامعه‌ی iام از j امین تکرار شبیه‌سازی است. همچنین واریانس از رابطه‌ی $\text{var}_i = \sum_{j=1}^{1000}(\hat{\theta}_{ij} - \hat{\theta}_i)^2 / 1000$ قابل برآورد است لذا برآوردی از MSE به صورت زیر برای طرح ناحیه‌ی کوچک محاسبه می‌شود:

$$\text{MSE}_i = \text{var}_i + \text{Bias}_i^2 \qquad i = 1, 2, \ldots, 16 \tag{۵}$$

در طرح دوم (شیوه‌ی کلاسیک)، رویکرد ناحیه‌ی کوچک وجود ندارد یعنی برای هر زیرجامعه، اندازه‌ی نمونه‌ی جداگانه تعیین و بر اساس آن، برآورد تفاضل درصدها محاسبه می‌شود ولی همچنان نمونه‌گیری با روش چرخشی اجرا می‌شود. با توجه به نااریب بودن برآوردگر تفاضل، از واریانس برآوردها یعنی $\text{var}_i = \frac{1}{1000}\sum_{j=1}^{j=1}(\hat{\theta}_{ij} - \hat{\theta}_i)^2$ به‌عنوان معیار

مقایسه استفاده خواهد شد. جدول ۳، این دو کمیت را برای سه میزان تداخل برای هر دو طرح (ناحیه‌ی کوچک و کلاسیک) ارائه کرده است.

جدول ۳– خطای برآوردگر تفاضل درصدها تحت سه میزان تداخل ۲۵، ۵۰ و ۷۵ درصد

ناحیه	طرح کلاسیک (واریانس)			طرح ناحیه‌ی کوچک MSE		
	٪۲۵	٪۵۰	٪۷۵	٪۲۵	٪۵۰	٪۷۵
۱	۳/۹۱۷	۳/۱۰۰	۲/۲۸۴	۷/۶۵۷	۶/۱۸۸	۵/۰۳۹
۲	۴/۴۱۷	۳/۴۹۱	۲/۵۶۴	۹/۴۱۸	۸/۶۵۶	۴/۰۵۸
۳	۴/۴۷۴	۳/۵۳۷	۲/۶۰۰	۱۰/۰۲۱	۸/۴۱۵	۶/۶۵۷
۴	۳/۹۴۸	۳/۱۲۴	۲/۳۰۰	۸/۷۷۹	۷/۵۲۰	۵/۳۷۹
۵	۳/۴۵۰	۲/۷۷۳	۲/۰۴۶	۱۱/۱۸۵	۹/۶۶۴	۷/۱۵۶
۶	۳/۹۳۳	۳/۱۱۴	۲/۲۹۵	۶/۸۷۸	۵/۸۲۱	۴/۱۰۹
۷	۳/۳۵۸	۲/۶۵۹	۲/۹۶۱	۹/۲۵۵	۶/۵۴۷	۵/۶۵۸
۸	۳/۴۱۴	۲/۷۰۸	۲/۰۰۴	۱۳/۶۵۴	۱۰/۳۵۴	۷/۶۵۷
۹	۳/۶۹۵	۲/۹۲۴	۲/۱۵۴	۸/۴۹۶	۷/۱۰۹	۵/۶۱۸
۱۰	۳/۳۲۵	۲/۶۳۸	۲/۹۵۲	۸/۹۴۸	۷/۹۶۶	۵/۰۸۱
۱۱	۳/۶۷۱	۲/۹۱۸	۲/۱۶۷	۷/۴۴۷	۶/۰۵۴	۳/۹۲۴
۱۲	۵/۰۹۰	۴/۰۱۹	۳/۹۴۸	۱۱/۶۵۱	۸/۹۰۲	۵/۰۲۸
۱۳	۴/۴۱۷	۳/۴۹۰	۲/۵۶۳	۱۱/۳۵۹	۸/۱۸۷	۵/۲۶۴
۱۴	۳/۱۴۳	۲/۵۰۲	۱/۸۶۰	۷/۶۵۸	۶/۸۱۸	۶/۰۳۸
۱۵	۳/۴۸۹	۲/۷۶۵	۲/۰۴۱	۸/۵۴۴	۶/۹۸۴	۶/۰۵۸
۱۶	۳/۳۵۰	۲/۶۵۷	۲/۹۶۳	۱۱/۹۶۲	۹/۲۸۰	۷/۳۱۹

۳- یافته‌ها و پیشنهادها

در این بخش می‌توان دو پرسش را مطرح کرد. اول، عملکرد طرح ناحیه‌ی کوچک در مقایسه با طرح کلاسیک چگونه است؟ در نگاه اول ممکن است ارقام جدول ۳، مأیوس‌کننده به نظر برسند یعنی طرح ناحیه‌ی کوچک را رقیبی جدی برای طرح کلاسیک نشان ندهند. با وجود این، هنگامی که به یاد می‌آوریم خطاهای طرح ناحیه‌ی کوچک حاصل ۶۷۱ نمونه ولی خطاهای طرح کلاسیک حاصل ۹۰۷۴ نمونه است به توانایی بالقوه طرح ناحیه‌ی کوچک پی می‌بریم.

پرسش دیگر به تأثیر میزان تداخل بر عملکرد دو طرح بر می‌گردد. همان‌طور که در جدول ۳ دیده می‌شود، افزایش میزان تداخل در کاهش خطای هر دو طرح مؤثر است ولی میزان کاهش خطا در طرح ناحیه‌ی کوچک بیش از طرح کلاسیک است. برای مثال، میزان کاهش خطا از تداخل ۲۵ درصد به تداخل ۷۵ درصد تحت طرح ناحیه‌ی کوچک برای زیرجامعه‌های ۸، ۱۲ و ۱۳ تقریباً به ۶ واحد می‌رسد در حالی که بیش‌ترین میزان کاهش خطا برای طرح کلاسیک، تنها ۲ واحد است.

این دو یافته که یکی به مسئله‌ی اندازه‌ی نمونه و دیگری به میزان تداخل اشاره دارد، راهکارهای زیر را برای بهبود طرح ناحیه‌ی کوچک یعنی کاهش خطای آن پیشنهاد می‌دهند.

- دسته‌بندی زیرجامعه‌ها و محاسبه‌ی اندازه‌ی نمونه‌ی جداگانه برای هر دسته. زیرجامعه‌های هر دسته باید نسبت به متغیر تحت مطالعه نسبتاً مشابه باشند. برای مثال، ممکن است بتوان ۱۶ زیرجامعه‌ی این شبیه‌سازی را بر اساس اطلاعات متغیر کمکی به سه دسته تفکیک کرد و برای برآورد تفاضل درصدهای هر دسته (و نه زیرجامعه‌های داخل آن)، اندازه‌ی نمونه را محاسبه کرد. سپس با استفاده از روش برآورد برای ناحیه‌ی کوچک و داده‌های نمونه‌ای موجود از هر دسته، برآورد تفاضل درصدهای هر زیرجامعه از آن دسته را به دست داد. انتظار می‌رود استفاده از داده‌های مربوط به زیرجامعه‌های مشابه منجر به برآوردهای دقیق‌تری شود.
- از آن‌جا که دسته‌بندی نیاز به اطلاعاتی به‌عنوان معیار دسته‌بندی دارد و اطلاعاتی تا این حد دقیق کمتر در دسترس است، افزایش اندازه‌ی نمونه‌ی کل (مقدار ۶۷۱) امکان دیگری است که در پیش رو است. به عبارت دیگر، با افزایش اندازه‌ی نمونه‌ی ۶۷۱تایی به دو برابر که همچنان نسبت به اندازه‌ی نمونه‌ی ۹۰۷۴تایی تفاوت چشمگیری دارد، داده‌های نمونه‌ای بیش‌تری را از زیرجامعه‌ها در اختیار می‌گذارد که می‌تواند به برآوردهای دقیق‌تری منجر شود.
- راه‌کار دیگر، یافتن یک میزان بهینه از تداخل است. تداخل‌های بالا به‌دلیل نیاز بیش‌تر به مراجعه به افراد یکسان که به افزایش بار پاسخگو[11] یا عدم دسترسی به آن‌ها منجر می‌شود، در اجرا با مشکل بی‌پاسخی روبه‌رو هستند. اگرچه در طرح

ناحیه‌ی کوچک اندازه‌ی نمونه در مقایسه با طرح کلاسیک بسیار کوچک‌تر است ولی باید تلاش شود تا همچنان میزان تداخل را پایین نگه داشت. انتظار می‌رود به‌کمک راهکارهای اندازه‌ی نمونه، تداخلی حدود ۳۰ درصد مناسب باشد (به [۲]، مراجعه شود).

این مقاله در صدد ارائه‌ی شیوه‌ای برای کاهش اندازه‌ی نمونه در آمارگیری‌های مکرر به‌ویژه آمارگیری‌های چرخشی بود زیرا تکرار این نوع از آمارگیری‌ها در فاصله‌های زمانی کوتاه مثلاً فصلی، به افزایش هزینه منجر می‌شود. یافته‌ها مؤید این امر هستند که استفاده از روش‌های برآورد برای ناحیه‌ی کوچک می‌توانند امکان کاهش اندازه‌ی نمونه را فراهم کنند ولی ضروری است پیش از استفاده از آن‌ها، راهکارهایی به خدمت گرفته شوند. شایان ذکر است که در این راهکارها هرگز تعداد نمونه به اندازه‌ای افزایش نمی‌یابد که ارزش اقتصادی استفاده از روش‌های برآورد ناحیه‌ی کوچک کم‌رنگ شود و همچنان این مزیت ارزنده خودنمایی می‌کند.

توضیحات

[1] Sample Surveys
[2] Repeated Surveys
[3] Cross-sectional Surveys
[4] Panel Suvey
[5] Rotation Survey
[6] Small Area Estimation Methods
[7] Model-based Estimators
[8] Design-based Estimators
[9] Mixed Model
[10] Delta Method
[11] Respondent Burden

مرجع‌ها

[۱] طاهری‌منزه، محمد (۱۳۸۰). استفاده از برآوردگر بیز تجربی برای برآورد نسبت بیکاری در سطح ناحیه‌های کوچک، پایان‌نامه‌ی کارشناسی ارشد، دانشگاه علامه طباطبایی.

[۲] نمازی‌راد، محمدرضا؛ نواب‌پور، حمیدرضا؛ خوشگویان‌فرد، علیرضا (۱۳۸۶). مقایسه‌ی آمارگیری‌های مکرر به کمک شبیه‌سازی، گزیده‌مطالب آماری، سال ۱۸، شماره‌ی ۱، صص ۳۵–۴۹.

[3] Firebaugh, G. (1997). Analyzing Repeated Surveys, Sage Publication.

[4] Khoshgooyanfard, A.-R. and Taheri Monazah, M. (2006). A Cost-Effective Strategy for Provincial Unemployment Estimation: A Small Area Approach, *Survey Methodology*, **32**, 105-114.