مجله گزیده‌مطالب آماری
سال ۱۸، شماره ۱، بهار و تابستان ۱۳۸۶، صص ۴۹-۳۵

مقایسه آمارگیری‌های مکرر به کمک شبیه‌سازی

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

[۱] پژوهشکده آمار
[۲] دانشگاه علامه طباطبایی
[۳] مرکز تحقیقات صدا و سیمای جمهوری اسلامی ایران

چکیده. این مقاله ابتدا به معرفی اجمالی آمارگیری‌های مکرر به‌ویژه آمارگیری چرخشی می‌پردازد. سپس، به کمک شبیه‌سازی، کارایی سه نوع آمارگیری مکرر– پانلی، مستقل و چرخشی– را در برآورد تغییرات پارامتری از نوع میانگین، در دو دوره متوالی مقایسه می‌کند. شبیه‌سازی در چهار حالت مختلف از همبستگی بین داده‌های دو دوره زمانی اجرا می‌شود و آمارگیری چرخشی نیز با ۵ میزان مختلف تداخل در شبیه‌سازی در نظر گرفته می‌شود. یافته‌ها، حاکی از آن هستند که آمارگیری پانلی دارای بیش‌ترین و آمارگیری مقطعی دارای کم‌ترین کارایی در برآورد تغییرات پارامتر مورد نظر است. همچنین، با افزایش همبستگی داده‌های دو دوره، بر کارایی برآورد تغییرات پارامتر در آمارگیری مکرر افزوده می‌شود.

۱- مقدمه

در بسیاری از نظام‌های آماری، برخی آمارگیری‌های موسوم به آمارگیری مکرر (Repeated Survey) یا آمارگیری در طول زمان (Survey Over Time)، به شکلی مستمر در دوره‌های زمانی منظم تکرار می‌شوند. برای این منظور، آمارگیری با شیوه

واژگان کلیدی: آمارگیری مکرر؛ آمارگیری پانلی؛ آمارگیری چرخشی؛ آمارگیری مقطعی مکرر؛ شبیه‌سازی؛ نمونه پایه؛ کارایی.
* نویسنده عهده‌دار مکاتبات

نمونه‌گیری و پرسشنامه « یکسان » در دوره‌هایی با فاصله زمانی برابر، مثلاً به‌طور فصلی، اجرا می‌شود. دو دلیل زیر را می‌توان از مهم‌ترین دلایل استفاده از این آمارگیری‌ها برشمرد.

اول: مطالعه یک متغیر در دوره‌های زمانی مختلف به شناسایی تغییرات و نوسانات آن در طول زمان کمک می‌کند. برای این منظور به شکلی مستمر در دوره‌های زمانی منظم برآوردهای مرتبط با آن متغیر محاسبه می‌شوند، تا الگو (Pattern) و روند (Trend) حاکم بر آن در طی زمان آشکار شود. برای مثال می‌توان نرخ مهاجرت از روستا به شهر را به‌منظور شناسایی روند تغییرات سالانه آن، هر ساله برآورد کرد تا پس از چند سال به روند تغییرات سالانه آن دست یافت. نکته اساسی در چنین مطالعاتی، آمارگیری برای چندین دوره متوالی و رعایت فاصله زمانی یکسان در آمارگیری‌ها است، تا امکان آشکار شدن روند تغییرات به شکل واقعی فراهم شود.

دوم: گاه هدف از آمارگیری، برآورد پارامتری در یک دوره زمانی خاص است، ولی به‌دلیل تغییراتی که برای این پارامتر در یک فاصله زمانی رخ می‌دهد، به اجبار آمارگیری در داخل این فاصله زمانی در چند نوبت تکرار می‌شود تا اثر دوره‌های زمانی مختلف حذف شود. برای مثال، اگر هدف یک آمارگیری، برآورد نرخ بیکاری در یک فاصله زمانی یک‌ساله باشد، نمی‌توان این آمارگیری را تنها در پایان سال اجرا کرد، زیرا نرخ حاصل از آن، نمایانگر نرخ بیکاری کل سال نیست. به عبارت دیگر، از آن‌جا که نرخ بیکاری در دوره‌های مختلف سال (مثلاً هر فصل) متفاوت است، نرخی که با این روش برآورد می‌شود، تنها نمایانگر نرخ بیکاری در زمان آمارگیری است. بنا بر این، آمارگیری در چند نوبت در سال (مثلاً هر فصل) اجرا و از ادغام برآوردهای دوره‌های مختلف، نرخ بیکاری سالانه، برآورد می‌شود. به عبارت دیگر، آمارگیری به‌جای آن‌که در یک مقطع زمانی خاص اجرا شود در « طول دوره مورد نظر » و در چند نوبت اجرا می‌شود.

همان‌طور که اشاره شد، برای تهیه داده‌های لازم برای تولید آماره‌ها در طول زمان، از آمارگیری‌های مکرر استفاده می‌شود. چنین آمارگیری‌هایی، اغلب به یکی از روش‌های سه‌گانه زیر اجرا می‌شوند:

آمارگیری مقطعی (Cross-sectional Survey): این روش، ساده‌ترین نوع آمارگیری در طول زمان است. در آمارگیری مقطعی در هر دوره زمانی (مثلاً هر فصل یا ماه)، نمونه جداگانه‌ای (مستقلی) را از جامعه تحت بررسی انتخاب می‌کنند. بنا بر این پس از چند دوره زمانی، چند نمونه « مستقل » از هم وجود خواهند داشت. بر اساس هر یک از این نمونه‌ها می‌توان پارامتر مورد نظر را برای آن دوره زمانی خاص برآورد کرد و در نهایت روند پارامتر مورد نظر را در طول زمان شناخت [۵].

آمارگیری پانلی (Panel Survey): در این روش، نمونه « ثابتی » در طول زمان تعقیب می‌شود. به عبارت دیگر به یک نمونه ثابت در دوره‌های زمانی مختلف مراجعه می‌شود. این بدان معنی است که از یک نمونه پس از چند دوره زمانی، چند مجموعه داده‌ها از دوره‌های زمانی مختلف وجود خواهند داشت. بنا بر این، در مقایسه با آمارگیری مقطعی، پس از چند دوره زمانی، چند نمونه « وابسته » وجود خواهد داشت. در آمارگیری پانلی، ثابت بودن نمونه در دوره‌های زمانی مختلف به حذف « اثر نمونه » در شناسایی دقیق‌تر تغییرات کمک می‌کند. به عبارت دیگر تفاوت مشاهده‌شده در دو دوره زمانی را نمی‌توان ناشی از تفاوت واحدهای نمونه‌ای قلمداد کرد. این نوع آمارگیری دقت برآورد تغییرات را افزایش می‌دهد [۲].

آمارگیری چرخشی (Rotation Survey): این روش آمارگیری، تلفیقی از دو روش آمارگیری مقطعی و پانلی است. در این روش، واحدهای نمونه‌ای در طول زمان نه کاملاً ثابت هستند و نه کاملاً متفاوت، بلکه بخشی از واحدهای نمونه‌ای بر اساس الگویی موسوم به الگوی چرخش(Rotation Design) تکرار می‌شوند. برای این منظور، ابتدا یک نمونه پایه (Master Sample) تعیین می‌شود. سپس، بر اساس الگوی چرخش از میان واحدهای نمونه پایه، تعداد نمونه لازم در هر دوره زمانی معین می‌شود. کسری از واحدهای نمونه هر دوره با واحدهای نمونه دوره قبل مشترک (متداخل) و بقیه متفاوت‌اند. بنا بر این، انتظار می‌رود که این نمونه، دارای اشتراکاتی (تداخل) با نمونه‌های دوره‌های دورتر نیز باشد [۳].

در ادامه، این روش آمارگیری در قالب یک مثال بررسی می‌شود. جدول ۱ مثال

ساده‌ای را از آمارگیری چرخشی ارائه می‌دهد. در این مثال، نمونه پایه، شامل ۱۱ واحد و نمونه لازم در هر دوره، شامل ۵ واحد است. همان‌طور که دیده می‌شود، چرخش واحدهای نمونه‌ای (ورود به نمونه و خروج از آن) برای شش دوره (تکرار) ارائه شده است.

در دوره اول، ۵ واحد اول از نمونه پایه انتخاب می‌شوند. در دوره دوم، دو واحد اول جای خود را به واحدهای ۶ و ۷ می‌دهند، و واحدهای ۳ تا ۵ همچنان در نمونه دوره دوم حضور دارند. در دوره سوم، واحدهای ۵، ۶ و ۷ از نمونه دوره دوم به نمونه دوره سوم منتقل می‌شوند و واحدهای ۸ و ۹ جایگزین واحدهای ۳ و ۴ می‌شوند و همین طور تا آخر. در پایان دوره چهارم هر یک از ۱۱ واحد نمونه پایه به‌گونه‌ای در یکی از دوره‌ها انتخاب شده‌اند. در این مثال ساده، نمونه هر دوره، با نمونه دوره قبل و دوره بعد از خود در سه واحد مشترک است (میزان تداخل $\frac{۳}{۵} \times ۱۰۰ = ۶۰$ درصد است). همچنین، نمونه‌های سه دوره متوالی تنها در یک واحد مشترک‌اند.

جدول ۱- مثالی از یک آمارگیری چرخشی

واحدهای	دوره زمانی					
نمونه پایه	اول	دوم	سوم	چهارم	پنجم	ششم
۱	√				√	√
۲	√				√	√
۳	√	√				√
۴	√	√				√
۵	√	√	√			
۶		√	√			
۷		√	√	√		
۸			√	√		
۹			√	√	√	
۱۰				√	√	
۱۱				√	√	√

تصمیم درباره این‌که برای اجرای آمارگیری مکرر از کدام یک از روش‌های سه‌گانه استفاده شود، به اهداف آمارگیری و ویژگی‌های این سه روش بستگی دارد. تعیین فاصله زمانی مناسب برای دوره‌های آمارگیری در هر سه روش اهمیت دارد. آمارگیری پانلی با مشکل دسترسی به واحدهای یکسان در تمام تکرارها و کاهش همکاری واحدهای نمونه‌ای به‌دلیل مراجعه‌های پی در پی یا تغییر نشانی روبه‌رو است (موسِر و کالتون ۱۹۷۹) و در عوض از مزیت « حذف اثر نمونه » برخوردار است (دانکَن و همکاران ۱۹۸۷). در آمارگیری مقطعی مشکل دسترسی وجود ندارد و در آمارگیری چرخشی نیز بر حسب میزان تداخل، این مشکل در مقایسه با آمارگیری پانلی کم‌رنگ‌تر است.

دو مقوله مهم در آمارگیری چرخشی تعیین نمونه پایه و الگوی چرخش است. محاسبه میانگین توان دوم خطای (Mean Squared Error) برآورد تغییرات پارامتر مورد نظر در آمارگیری چرخشی به‌دلیل تداخل نمونه‌ها، دشوارتر از آمارگیری مقطعی (که تغییرات را بر اساس دو نمونه مستقل برآورد می‌کند) و آمارگیری پانلی (که تغییرات را بر اساس دو نمونه کاملاً وابسته برآورد می‌کند) است [۴].

در قسمت‌های آتیِ این مقاله، ابتدا در بخش ۲ به بررسی کامل شبیه‌سازی طرح‌ریزی‌شده در پژوهش حاضر پرداخته می‌شود و سپس در بخش ۳ نتایج نهایی این پژوهش به همراه چند پیشنهاد آورده شده است.

۲- شبیه‌سازی

در این بخش، از شبیه‌سازی مونت کارلو برای مقایسه سه روش معرفی‌شده، استفاده می‌شود. برای این منظور، فرض می‌شود هدف آمارگیری، برآورد تغییرات میانگین متغیر X از یک جامعه با جمعیت ۲۰۰,۰۰۰ واحد در دو دوره زمانی متوالی است. برای آن‌که آمارگیری مکرر توجیه‌پذیر باشد، توزیع متغیر X در دو دوره متوالی، متفاوت در نظر گرفته شده و داده‌های این دو دوره به‌گونه‌ای تولید می‌شوند که دارای همبستگی باشند.

۱-۲- طرح مطالعه

برای انجام مقایسه، پس از محاسبه اندازه نمونه برای هر دوره و انتخاب بزرگ‌ترین اندازه نمونه به‌عنوان اندازه نمونه نهایی به‌منظور حفظ دقت مورد نظر، انتخاب نمونه‌ها به سه روش صورت می‌گیرد. در حالت نخست، نمونه‌های مستقلی از هر دوره فرضی انتخاب می‌شوند. در حالت دوم، مقادیر نمونه‌های دوره دوم مربوط به همان « واحدهای » نمونه دوره اول است، به‌طوری که یک مجموعه داده‌های پانلی (Panel Data) به دست آید. حالت سوم به نمونه‌گیری چرخشی اختصاص دارد. برای آن‌که اثر میزان تداخل نیز ارزیابی شود، نمونه‌گیری چرخشی، در پنج وضعیت مختلف (تداخل‌های ۱۰، ۲۰، ۳۰، ۴۰ و ۵۰ درصد) در نظر گرفته می‌شود. برای مثال، در آمارگیری چرخشی با تداخل ۱۰ درصد، ۱۰ درصد از مقادیر نمونه‌های دوره دوم از واحدهای نمونه دوره اول به دست می‌آیند و بقیه واحدها متفاوت (مستقل) از واحدهای دوره اول به دست می‌آیند. بنا بر این، ۷ حالت (نمونه‌گیری مقطعی، نمونه‌گیری پانلی و نمونه‌گیری چرخشی با پنج درصد مختلف تداخل) وجود دارد. شایان ذکر است، یکسان بودن واحدهای نمونه‌ای در دو دوره در نمونه پانلی یا چرخشی، به معنای برابری مقدار متغیر مورد نظر در دو دوره نیست، زیرا جامعه‌های آماری در دو دوره زمانی از توزیعی با مقادیر مختلف پارامترها تولید می‌شوند که با یکدیگر همبسته‌اند.

از آن‌جا که انتظار می‌رود کارایی سه روش آمارگیری، تحت تأثیر میزان همبستگی داده‌های دو دوره باشد، شبیه‌سازی تحت چهار مقدار (گردشده) از همبستگی ۰/۲، ۰/۴، ۰/۶ و ۰/۸ تکرار شده است تا بدین وسیله، اثر میزان همبستگی بر کارایی، سنجیده شود. در مجموع، از ترکیب هفت حالت آمارگیری که هر یک در چهار وضعیت همبستگی شبیه‌سازی شده‌اند، به ۲۸ حالت مختلف شبیه‌سازی می‌رسیم، که پاسخگویی به پرسش‌های زیر را ممکن می‌سازند:

- کدام یک از سه روش آمارگیری مقطعی، پانلی و چرخشی کاراتر است؟
- آیا افزایش میزان تداخل بر دقت آمارگیری چرخشی اثر می‌گذارد؟
- افزایش همبستگی بین داده‌های دو دوره متوالی چه اثری در کارایی سه روش

آمارگیری دارد؟

- آیا افزایش همبستگی بین داده‌های دو دوره متوالی می‌تواند باعث شود که با تداخل کم‌تری به کارایی بیش‌تری دست یابیم؟

۲-۲- تولید داده‌ها

برای تولید داده‌های مورد نیاز در مطالعه حاضر، از نرم‌افزار SPSS استفاده شده است. در گام اول، ۲۰۰,۰۰۰ داده شبه‌تصادفی از یک توزیع نرمال با میانگین ۱۰۰۰ و انحراف معیار ۱۰۰ تولید شد. این اعداد در کل مراحل شبیه‌سازی به‌عنوان جامعه آماری در دوره زمانی اول در نظر گرفته می‌شوند. در گام دوم، به‌منظور تولید داده‌های مربوط به جامعه آماری در دوره زمانی دوم، از ترکیب‌های خطی اعداد به دست آمده در گام اول و یک متغیر تصادفی نرمال در نرم‌افزار مذکور استفاده شد تا همبستگی مورد نیاز بین داده‌های مربوط به دو دوره زمانی حاصل شود. برای مثال برای تولید جامعه آماری در دوره دوم که با داده‌های دوره اول دارای همبستگی ۰/۶ باشد به شکل زیر عمل شد:

$$X_۲ = \frac{۳}{۲} \times X_۱ + RV.\ Normal(۵۰۰,\ ۲۰۰)$$

که در آن تابع $RV.\ Normal$ داده‌های شبه‌تصادفی از توزیع نرمال با میانگین ۵۰۰ و انحراف معیار ۲۰۰ تولید می‌کند. پس جامعه آماری دوره زمانی دوم دارای توزیع نرمال با میانگین ۲۰۰۰ و انحراف معیار ۲۵۰ است. بر این اساس، مشخصات جامعه‌های آماری برای چهار مقدار همبستگی به شرح زیر است:

۱: $X_۱ \sim N(۱۰۰۰\ ,\ ۱۰۰),\quad X_۲ \sim N(۱۹۵۰\ ,\ ۵۷۰),\quad r_{X_۱ r_{X_۲}} = ۰/۲۱$

۲: $X_۱ \sim N(۱۰۰۰\ ,\ ۱۰۰),\quad X_۲ \sim N(۱۶۵۰\ ,\ ۲۹۵),\quad r_{X_۱ r_{X_۲}} = ۰/۴۰۲$

۳: $X_۱ \sim N(۱۰۰۰\ ,\ ۱۰۰),\quad X_۲ \sim N(۲۰۰۰\ ,\ ۲۵۰),\quad r_{X_۱ r_{X_۲}} = ۰/۶۰۱$

۴: $X_۱ \sim N(۱۰۰۰\ ,\ ۱۰۰),\quad X_۲ \sim N(۱۶۰۰\ ,\ ۱۹۰),\quad r_{X_۱ r_{X_۲}} = ۰/۸$

که در آن $X_۱$ و $X_۲$ به‌ترتیب متغیرهای موردنظر در دوره‌های زمانی اول و دوم هستند.

برای تعیین اندازه نمونه، جامعه آماری حالت نخست (همبستگی ۰/۲) مبنا قرار گرفت

زیرا بیش‌ترین انحراف معیار (مقدار ۵۷۰) که منجر به بیش‌ترین اندازه نمونه نیز می‌شود، به جامعه آماری دوره زمانی دوم تعلق دارد. اندازه نمونه از فرمول نمونه‌گیری تصادفی ساده بدون جایگذاری با سطح اطمینان ۹۵ درصد و حاشیه‌های خطای (Margin of Error) ۴۵/۵۱ و ۱۰ به‌ترتیب برای دوره زمانی دوم و اول برابر با ۶۰۱ و ۳۸۴ محاسبه شد. در نهایت بیش‌ترین این دو اندازه نمونه یعنی ۶۰۱ مورد استفاده قرار گرفت تا خطای مورد نظر برای هر دو دوره زمانی حفظ شود.

برای هر یک از حالات ۲۸گانه پیش، ۱۰۰۰ نمونه ۶۰۱ تایی انتخاب شد تا بر اساس آن، مقایسه کارایی روش‌ها صورت گیرد. برای مثال، ۱۰۰۰ زوج نمونه ۶۰۱ تایی پانلی، ۱۰۰۰ زوج نمونه ۶۰۱ تایی چرخشی با تداخل ۱۰ درصد و... برای برآورد تغییرات میانگین دو دوره زمانی متوالی انتخاب شدند.

۳-۲- معیارهای مقایسه

به‌منظور بررسی نتایج به دست آمده در شبیه‌سازی، با توجه به ناریب بودن برآوردگرها (میانگین معمولی برآوردگر ناریبی در نمونه‌گیری تصادفی ساده است) از برآوردگر واریانس که به شکل زیر تعریف می‌شود، به‌عنوان معیار مقایسه روش‌های آمارگیری استفاده شد.

$$Var = \frac{\sum_{i=1}^{1000} (\hat{\theta}_i - \hat{\theta})^2}{1000} \qquad (1)$$

که در آن:

$$\hat{\theta}_i = \hat{\bar{X}}_{2i} - \hat{\bar{X}}_{1i} \qquad i = 1, 2, \ldots, 1000$$
$$\hat{\theta} = \hat{\bar{X}}_2 - \hat{\bar{X}}_1$$

که در رابطه‌های بالا $\hat{\bar{X}}_{1i}$ و $\hat{\bar{X}}_{2i}$ به‌ترتیب میانگین‌های تکرار iام برای دوره‌های اول و دوم و $\hat{\bar{X}}_1$ و $\hat{\bar{X}}_2$ به‌ترتیب نشان‌دهنده میانگین‌های ۱۰۰۰ نمونه به دست آمده از دوره‌های زمانی اول و دوم هستند.

پس از محاسبه واریانس برای هر یک از حالات ۲۸گانه شبیه‌سازی، به‌منظور بررسی بیش‌تر کارایی آمارگیری چرخشی و پانلی، برآورد کارایی نسبی مجانبی (Estimated Asymptomatic Relative Efficiency) از رابطه زیر محاسبه شد. به عبارت دیگر، آمارگیری‌های متداخل (چرخشی و پانلی) با آمارگیری مقطعی مقایسه شدند (آمارگیری پانلی را نیز می‌توان نوعی آمارگیری چرخشی با تداخل ۱۰۰ درصد قلمداد کرد).

$$EARE_i = \frac{Var_{ij}^{\text{Overlapping}}}{Var_i^{\text{Independent}}};\quad i = ۱, ۲, ۳, ۴;\quad j = ۱, ۲, ۳, ۴, ۵ \qquad (۲)$$

که در این رابطه، $Var_{ij}^{\text{Overlapping}}$ نشان‌دهنده واریانس روش آمارگیری متداخل برای همبستگی iام و تداخل jام و $Var_i^{\text{Independent}}$ نشان‌دهنده واریانس روش آمارگیری مقطعی برای همبستگی iام است. بنا بر این، $EARE_i$ با مقداری کمتر از « یک » بر کاراتر بودن روش آمارگیری متداخل دلالت دارد.

همچنین، برای بررسی اثر همبستگی، واریانس هر روش با همبستگی‌های ۰/۴، ۰/۶ و ۰/۸ با واریانس همان روش هنگامی که همبستگی ۰/۲ است با رابطه زیر مقایسه شد:

$$EARV_i = \frac{Var_i^{C_j}}{Var_i^{C_۱}};\quad i = ۱, ۲, \ldots, ۷;\quad j = ۲, ۳, ۴ \qquad (۳)$$

$$c_۱ = ۰/۲ ,\ c_۲ = ۰/۴ ,\ c_۳ = ۰/۶ ,\ c_۴ = ۰/۸$$

که در این رابطه، $Var_i^{C_۱}$ در مخرج کسر، نشان‌دهنده واریانس روش‌های مختلف آمارگیری با همبستگی ۰/۲ و تداخل iام $(i = ۱, ۲, \ldots, ۷)$ و $Var_i^{C_j}$ $(j = ۲, ۳, ۴)$ نشان‌دهنده واریانس روش‌های مختلف آمارگیری برای همبستگی‌های ۰/۴، ۰/۶ و ۰/۸ است. بنا بر این، $EARV_i$ با مقداری کمتر از « یک » نشان می‌دهد که افزایش همبستگی میان داده‌های دو دوره زمانی، افزایش کارایی در نمونه‌گیری‌های مکرر را سبب می‌شود.

۴-۲- یافته‌ها

جدول ۲ مقادیر واریانس را بر اساس رابطه (۱) برای هر یک از حالت‌های ۲۸گانه مطالعه شبیه‌سازی نشان می‌دهد. بر این اساس می‌توان به چهار پرسش بخش (۱-۲) پاسخ داد. همان‌طور که دیده می‌شود آمارگیری پانلی همواره دارای کم‌ترین واریانس و آمارگیری مقطعی همواره دارای بیش‌ترین واریانس در برآورد تغییرات میانگین است. بنا بر این، دور از انتظار نیست که آمارگیری چرخشی وضعیتی بین این دو داشته باشد.

جدول ۲- برآورد واریانس تفاضل میانگین‌ها در دو دوره زمانی متوالی برای هر یک از حالت‌های ۲۸گانه مطالعه شبیه‌سازی

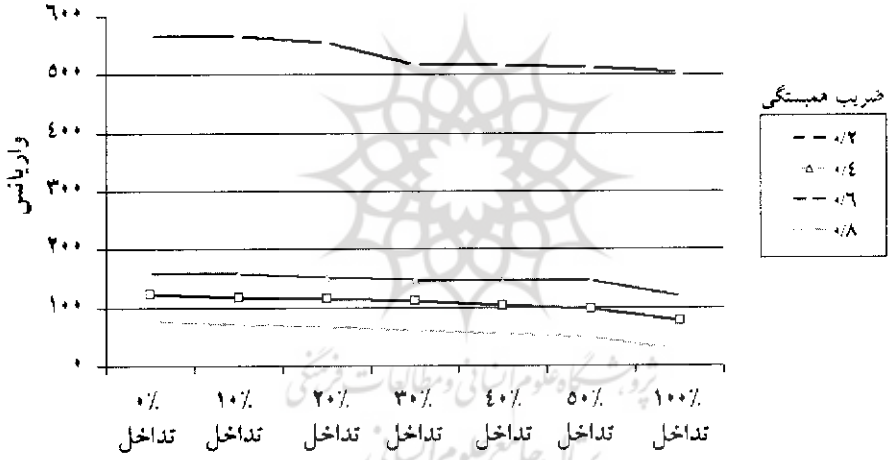
روش آمارگیری		همبستگی بین داده‌های دو دوره متوالی			
		۰/۲	۰/۴	۰/۶	۰/۸
آمارگیری مقطعی (تداخل ۰٪)		۵۶۷	۱۶۰	۱۲۳	۷۸
آمارگیری پانلی (تداخل ۱۰۰٪)		۵۰۵	۱۲۰	۷۸	۲۶
آمارگیری چرخشی	تداخل ۱۰٪	۵۶۶	۱۵۸	۱۱۸	۷۱
	تداخل ۲۰٪	۵۵۵	۱۵۳	۱۱۵	۶۷
	تداخل ۳۰٪	۵۱۷	۱۴۷	۱۱۱	۶۰
	تداخل ۴۰٪	۵۱۶	۱۴۶	۱۰۳	۵۵
	تداخل ۵۰٪	۵۱۲	۱۴۵	۹۸	۴۹

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

با افزایش درصد تداخل، همواره از مقدار واریانس آمارگیری چرخشی کاسته می‌شود. در صورت وجود همبستگی متوسط و قوی بین داده‌های دو دوره زمانی متوالی، تفاوت زیادی بین برآورد واریانس تفاضل میانگین‌ها برای تداخل‌های مختلف دیده نمی‌شود. این در حالی است که در صورت وجود همبستگی ضعیف بین داده‌های دو دوره، شکافی بین

برآورد واریانس‌ها برای تداخل‌های ۱۰ و ۲۰ درصد با سه تداخل دیگر به چشم می‌خورد. به‌عنوان مثال، برای همبستگی ۰/۲ برآورد واریانس برای دو تداخل ۱۰ و ۲۰ درصد به یکدیگر نزدیک ولی از برآورد واریانس‌ها برای تداخل‌های ۳۰، ۴۰ و ۵۰ درصد دور هستند.

در مجموع می‌توان گفت برای همبستگی متوسط بین داده‌های دو دوره، دست کم تداخل لازم برای آن‌که کارایی آمارگیری چرخشی، اختلاف قابل توجهی با آمارگیری مقطعی داشته باشد، مقداری بین ۲۰ تا ۳۰ درصد است. با توجه به این نکته که معمولاً میزان همبستگی داده‌های دو دوره زمانی متوالی نامعلوم است، استفاده از آمارگیری چرخشی با این درصد از تداخل پیشنهاد می‌شود.



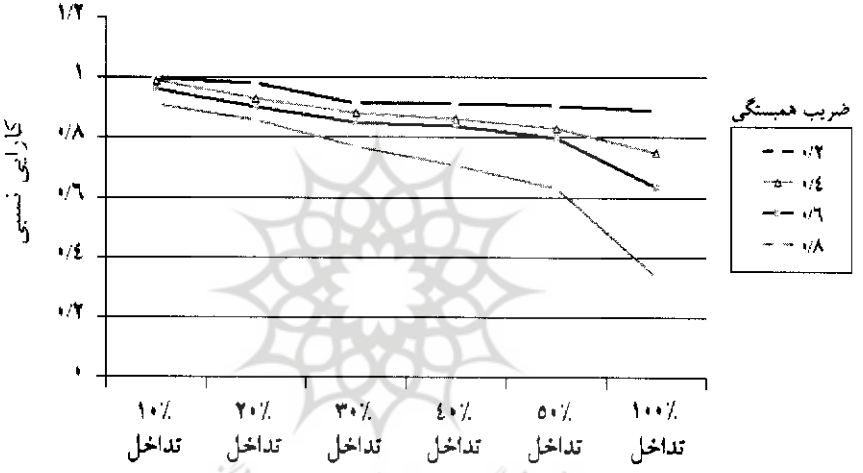
شکل ۱- مقایسه واریانس روش‌های مختلف آمارگیری مکرر با در نظر گرفتن مقادیر متفاوت ضریب همبستگی

شکل ۱ واریانس را برای روش‌های مختلف آمارگیری با در نظر گرفتن مقادیر مختلف ضریب همبستگی نشان می‌دهد.

اگرچه با افزایش همبستگی از مقدار واریانس برای هر سه روش آمارگیری کاسته می‌شود، ولی اختلاف واریانس مربوط به همبستگی ۰/۲ با واریانس سایر همبستگی‌ها چشمگیرتر است. به عبارت دیگر، وجود یک همبستگی نسبتاً متوسط می‌تواند کارایی قابل

قبولی را برای برآوردها تضمین کند. این امر در آمارگیری چرخشی بسیار چشمگیر است، به‌طوری که برای همبستگی ۰/۴ و بیش‌تر، برآورد واریانس تفاضل دو میانگین حداکثر به ۱۵۸ می‌رسد (تداخل ۱۰ درصد) در حالی که برای همبستگی ۰/۲، کم‌ترین برآورد برای واریانس برابر با ۵۱۲ است (تداخل ۵۰ درصد). برای همبستگی ۰/۸ برآورد واریانس حتی برای تداخل ۱۰ درصد نیز دو رقمی است.

شکل ۲ برآوردهای کارایی نسبی مجانبی را برای روش آمارگیری متداخل نسبت به آمارگیری مقطعی با استفاده از رابطه (۲) نشان می‌دهد. می‌توان مشاهده کرد که با هر

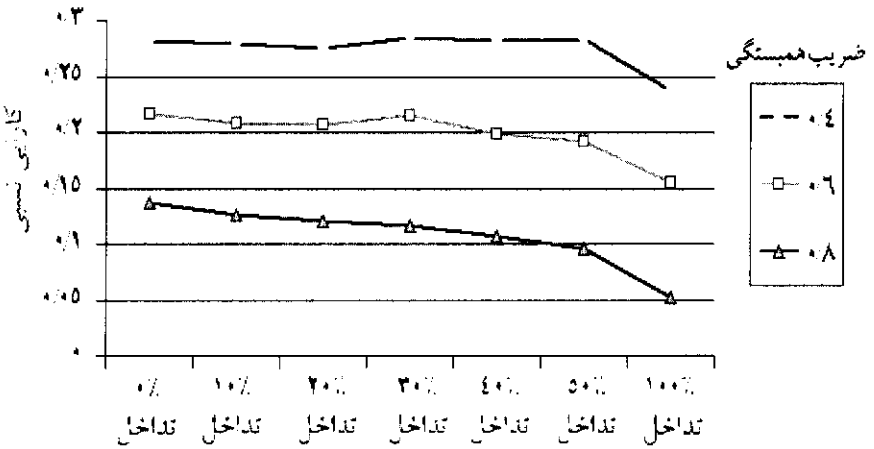


شکل ۲- مقایسه کارایی مجانبی نمونه‌گیری چرخشی و پانلی نسبت به نمونه‌گیری مقطعی با در نظر گرفتن مقادیر مختلف ضریب همبستگی

میزان همبستگی بین داده‌های دو دوره زمانی، با افزایش میزان تداخل، کارایی نمونه‌گیری متداخل نسبت به نمونه‌گیری مقطعی بیش‌تر می‌شود.

با بررسی شکل‌های ۱ و ۲ می‌توان نتیجه گرفت که افزایش تداخل و افزایش همبستگی هر یک به نوبه خود باعث افزایش کارایی می‌شوند.

شکل ۳ با استفاده از رابطه (۳) برآورد واریانس تفاضل میانگین‌های دو دوره زمانی متوالی را برای روش‌های مختلف آمارگیری با همبستگی‌های ۰/۴، ۰/۶ و ۰/۸ نسبت به همبستگی ۰/۲ نشان می‌دهد.



شکل ۳- مقایسه واریانس‌های روش‌های مختلف نمونه‌گیری مکرر با همبستگی‌های ۰/۴ ، ۰/۶ و ۰/۸ با واریانس همان روش‌ها به همبستگی ۰/۲

نمودار ۳ نشان می‌دهد که با افزایش میزان همبستگی در داده‌های دو دوره زمانی متوالی، برآورد واریانس تفاضل دو میانگین در همه روش‌های نمونه‌گیری در طی زمان کاهش می‌یابد.

۳- نتیجه‌گیری و پیشنهاد

با بررسی یافته‌های حاصل از مطالعه شبیه‌سازی، می‌توان به سؤال‌های مطرح‌شده در بخش ۲ پاسخ داد. نتایج در جدول ۲ نشان می‌دهند که با افزایش میزان تداخل، از میزان خطای برآورد تفاضل میانگین صفت مورد بررسی در دو دوره زمانی متوالی کاسته می‌شود. بنا بر این، اولاً نمونه‌گیری متداخل در مواردی که برآورد تغییرات مد نظر باشد، کاراتر از نمونه‌گیری مقطعی عمل می‌کند و ثانیاً افزایش تداخل در نمونه‌گیری نقش به‌سزایی در کاهش خطای نمونه‌گیری ایفا می‌کند. شکل‌های ۲ و ۳ بیانگر این مطلب هستند که افزایش همبستگی بین داده‌های دو دوره زمانی متوالی می‌تواند باعث افزایش کارایی در برآوردها شود. بنا بر این، انتظار می‌رود در مواقعی که داده‌های دو دوره زمانی متوالی

دارای همبستگی متوسطی باشند، بتوان با استفاده از تداخل کم‌تر به کارایی بیش‌تری دست پیدا کرد. اما از آن‌جا که غالباً همبستگی بین داده‌های دو دوره نامعلوم است، می‌توان با به‌کارگیری نمونه‌گیری چرخشی با تداخلی بین ۲۰ تا ۳۰ درصد به کارایی مناسبی دست یافت.

با وجود این، مراجعه به « داده‌های واقعی » حاصل از طرح‌هایی که در سال‌های گذشته توسط مرکز آمار ایران اجرا شده است، می‌تواند مبنایی برای مطالعه همبستگی بین داده‌های دو دوره زمانی متوالی فراهم کند تا به این ترتیب، به درصد تداخل واقع‌بینانه‌تری دست یابیم.

نکته قابل توجه دیگر که در مقدمه مقاله نیز به آن اشاره شد با آمارگیری پانلی و تا حدی آمارگیری چرخشی ارتباط دارد. یافته‌ها برتری آمارگیری پانلی را در تمام شرایط شبیه‌سازی نشان دادند. همچنین، آمارگیری‌های چرخشی با تداخل بالا همواره کاراتر از آمارگیری‌های چرخشی با تداخل پایین بودند. اگرچه این امر از بعد نظری تأیید شد ولی در عمل، مراجعه به نمونه‌ای ثابت در چندین دوره زمانی از مشارکت پاسخگویان به‌دلیل بار پاسخگویی (Response Burden) می‌کاهد یا دسترسی به آن‌ها را ناممکن می‌سازد که این دو به افزایش نرخ بی‌پاسخی (Non-response Rate) منجر خواهند شد. به عبارت دیگر، طولانی بودن آمارگیری‌های مکرر، آمارگیری پانلی یا آمارگیری چرخشی با تداخل بالا را با کاهش تعداد پاسخگویان روبه‌رو خواهد کرد. بنا بر این انتظار می‌رود از کارایی آمارگیری پانلی یا آمارگیری چرخشی با تداخل بالا کاسته شود. این امر، دوباره بر انتخاب درصدی از تداخل که واقع‌بینانه باشد و آمارگیری را در تکرارهای دورتر آن با مشکل روبه‌رو نسازد، تأکید می‌کند.

سپاس‌گزاری

پژوهش انجام‌شده از حمایت‌های مالی و معنوی پژوهشکده آمار برخوردار بوده است.

مرجع‌ها

[1] Duncan, G.J.; Juster, F.T.; Morgan, J.N. (1987), The Role of Panel Studies In Research on Economic Behavior, *Transportation Research,* 21a (4/5), 249-263.

[2] Duncan, G. and Kalton, G. (1987), "Issues of design and analysis of surveys across time", International Statistical Review 55, 97-117.

[3] Eckler, A.R. (1955), Rotation Sampling, *Annals of Mathematical Statistics* **26**, 664-685.

[4] Ghangurde, P.D.; Rao, J.N.K. (1969), Some Results on Sampling over Two Occasions, *Sankhya*, **3**, 463-472.

[5] Menard, S. (1991), Longitudinal Research, Sage University Paper series on Quantitative Applications in the Social Science, 07-076. Newbury Park, CS: Sage.

[6] Moser, C.A.; Kalton, G. (1979), *Survey Methods in Social Investigation, 2nd Edition*, Heinemann Educational Books: London.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی