



مقایسه‌ی روش‌های برآورد برای ناحیه‌های کوچک در برآورد نرخ بی‌کاری

آرمان بیداربخت‌نیا و روشنگر علی‌اکبری صبا*

مرکز آمار ایران

چکیده. نیاز به ارائه‌ی برآوردهایی برای ناحیه‌های کوچک به‌خصوص در طرح‌های خانواری مرکز آمار ایران، با توجه به بار پاسخگو و هزینه‌ی سنگین این طرح‌ها به‌سرعت رو به رشد است. اما چالش اساسی مرکز آمار ایران در استفاده از این روش‌های سودمند، عدم دسترسی به متغیرهای کمکی مناسب در فاصله‌ی بین دو سرشماری عمومی نفوس و مسکن است. در این مقاله با استفاده از نمونه‌های تکراری که از فایل سرشماری عمومی نفوس و مسکن ۱۳۷۵ انتخاب شده است، ضمن بررسی چند روش شناخته‌شده‌ی برآورد برای ناحیه‌های کوچک و تأثیر کهنگی اطلاعات کمکی بر هر یک از آن‌ها، پیشنهاد می‌شود که به‌منظور استفاده‌ی بهینه از این روش‌ها یک مقدار ثابت به‌عنوان مینیمم اندازه‌ی نمونه در ناحیه‌های کوچک در نظر گرفته شود. همچنین با استفاده از معیارهای مختلف، روش‌های برگزیده از بین روش‌های مورد بررسی معرفی می‌شوند.

واژگان کلیدی. برآوردهای مبتنی بر طرح؛ برآورد برای ناحیه‌ی کوچک؛ نرخ بی‌کاری؛ پساطبقه.

* نویسنده‌ی عهده‌دار مکاتبات

۱ مقدمه

آمارگیری‌های نمونه‌ای نه تنها برای برآورد پارامترهای جامعه، بلکه در بیش‌تر مواقع به منظور برآورد پارامترهای مورد نظر برای زیرجامعه‌ها به کار می‌روند. زیرجامعه‌ها می‌توانند گروه‌های جغرافیایی یا جمعیتی باشند که اصطلاحاً به عنوان حوزه‌ها یا ناحیه‌های کوچک از آن‌ها یاد می‌شود. آن‌چه مسلم است، این‌که همواره محدودیت منابع در مراکز تولید آمار وجود داشته و این مسئله منجر به استفاده از روش‌هایی برای کاهش هزینه‌ی آمارگیری‌ها شده است. از آن‌جا که بهینه کردن طرح نمونه‌گیری در سطح ناحیه‌های کوچک، باعث افزایش قابل توجه هزینه‌ی آمارگیری می‌شود، آمارشناسان علاقه‌مند هستند با نمونه‌ی بهینه‌شده برای کل جامعه، بتوانند در سطوح کوچک‌تر، برآوردهایی معتبر برای پارامترهای مورد نظر به دست آورند. روش‌های برآورد برای ناحیه‌ی کوچک، روش‌هایی هستند که با استفاده از مقادیر متغیرهای کمکی در سطح ناحیه‌های کوچک، برآورد برای این ناحیه‌ها را بدون نیاز به افزایش تعداد نمونه فراهم می‌کنند؛ ضمن این‌که این روش‌ها باعث کاهش خطاهای غیر نمونه‌گیری نیز می‌شوند.

در حال حاضر با توجه به تنوع روش‌های برآورد برای ناحیه‌های کوچک و همچنین عدم دستیابی به اطلاعات کمکی معتبر در فاصله‌ی ده‌ساله‌ی بین دو سرشماری عمومی نفوس و مسکن، انتخاب روش مناسب و همچنین متغیر کمکی مناسب برای محاسبه‌ی برآوردها از جمله‌ی چالش‌هایی است که مرکز آمار ایران در استفاده از این روش سودمند در طرح نیروی کار و همچنین سایر طرح‌های آمارگیری با آن مواجه است.

در این مقاله سعی شده است با استفاده از نمونه‌های تکراری که از داده‌های مربوط به سرشماری عمومی نفوس و مسکن ۱۳۷۵ استخراج شده‌اند و با به‌کارگیری اطلاعات کمکی از سرشماری عمومی نفوس و مسکن ۱۳۶۵، برخی از روش‌های برآورد برای ناحیه‌های کوچک برای محاسبه‌ی برآورد نرخ بی‌کاری به تفکیک استان‌ها با یکدیگر مقایسه شوند. بدین منظور، نخست در بخش ۱ روش‌های برآورد مورد استفاده در این مقاله برای ناحیه‌های کوچک معرفی می‌شوند. در بخش ۲ به برخی نکات عملی و محاسباتی اشاره می‌شود. در بخش ۳ نیز علاوه بر معرفی اندازه‌های کیفیت، نتایج به دست آمده از مقایسه‌ی نمودارها شرح داده می‌شوند.

۲ روش‌شناسی

به‌طورکلی برآوردهای مورد استفاده برای ناحیه‌های کوچک را می‌توان به سه دسته تقسیم کرد (راو، ۲۰۰۳):

آ) برآوردگرهای مستقیم،

ب) برآوردگرهای غیر مستقیم،

پ) برآوردگرهای ترکیبی.

برآوردگرهای مستقیم، برآوردگرهایی هستند که برای برآورد پارامتر در یک ناحیه ی کوچک، فقط از داده های همان ناحیه استفاده می کنند و معمولاً برآوردهایی نااریب به دست می دهند که به دلیل کوچک بودن این ناحیه ها و نیز اندازه ی کم نمونه ها در ناحیه های کوچک، دارای تغییرپذیری زیادی هستند. برآوردگر رگرسیونی برای ناحیه های کوچک نیز از این نوع برآوردگرها است که در صورت بزرگ بودن اندازه ی نمونه در این ناحیه ها، بسیار خوب عمل می کند.

برآوردگرهای غیر مستقیم برای محاسبه ی برآوردها در سطح ناحیه های کوچک، علاوه بر مشاهدات این ناحیه ها از مشاهدات ناحیه های دیگر نیز استفاده می کنند و به همین دلیل، دقت بیشتری دارند. برآوردگرهای به کار برده شده برای ناحیه های کوچک در این مقاله، از نوع برآوردگرهای غیر مستقیم یا ترکیب آن ها با برآوردگرهای مستقیم اند.

۲/۱ برآوردگر ترکیبی

بر اساس تعریف راتو (۲۰۰۳) اگر با استفاده از برآورد مستقیم معتبر برای ناحیه ی بزرگی که چند ناحیه ی کوچک را پوشش می دهد، تحت فرض همگن بودن آن ناحیه های کوچک، برآوردهای غیر مستقیم برای ناحیه های کوچک به دست آوریم، به برآوردهای حاصل، برآورد ترکیبی، و به برآوردگر متناظر، برآوردگر ترکیبی گفته می شود، که برآوردگری اریب است.

فرض کنید جامعه ی U به I ناحیه ی کوچک تقسیم شود. هدف، برآورد مقدار کل مشخصه ی Y در ناحیه ی کوچک i ام است. اگر جامعه را با استفاده از اطلاعات متغیری کمکی به G طبقه ی همگن تقسیم کنیم، به گونه ای که تعداد واحدهای آماری در سطح ناحیه ی کوچک i ام ($i = 1, \dots, I$) برای پساتبقه ی g ام ($g = 1, \dots, G$) با استفاده از آمارگیری های معتبر یا از سرشماری معلوم باشد، برآورد ترکیبی برای مقدار کل مشخصه ی Y در ناحیه ی کوچک i ام به شکل

$$(۱) \quad \hat{Y}_{i,SY} = \sum_g N_{ig} \frac{\hat{Y}_{\cdot g}}{\hat{N}_{\cdot g}}$$

محاسبه می شود، که در آن، $\hat{Y}_{\cdot g}$ برآورد مستقیم مقدار کل مشخصه ی Y در پساتبقه ی g ام، N_{ig} تعداد برای ناحیه ی کوچک i ام و پساتبقه ی g ام، و $\hat{N}_{\cdot g}$ برآورد تعداد کل در پساتبقه ی g ام است.

اگر مشخصه‌ی Y یک متغیر دوحالتی باشد، یعنی $y_j = ۰$ یا $y_j = ۱$ ، $(j = ۱, \dots, n)$ ، برآورد ترکیبی برای نسبت p_i در ناحیه‌ی کوچک i ام از

$$(۲) \quad \hat{p}_{i,S_Y} = \frac{\sum_g N_{ig} \hat{p}_{\cdot g}}{\sum_g N_{ig}}$$

به دست می‌آید، که مقدار $\hat{p}_{\cdot g}$ برآورد مستقیم نسبت در پساتبقه‌ی g ام، و $\sum_g N_{ig} = N_i$ تعداد کل در ناحیه‌ی کوچک i ام است، که از سرشماری یا آمارگیری معتبر به دست آمده است.

با تغییر $\hat{N}_{\cdot g}$ به $N_{\cdot g}$ در رابطه‌ی (۱) و تقسیم طرفین رابطه بر تعداد کل واحدها در ناحیه‌ی کوچک i ام، یک برآوردگر ترکیبی جایگزین به صورت

$$(۳) \quad \hat{p}_{i,A1} = \frac{\sum_g N_{ig} (\hat{Y}_{\cdot g} / N_{\cdot g})}{\sum_g N_{ig}}$$

به دست می‌آید، که $\hat{Y}_{\cdot g}$ عبارت است از برآورد مستقیم مقدار کل برای مشخصه‌ی مورد بررسی در پساتبقه‌ی g ام.

۲/۲ برآوردگر مرکب

یک روش برای برقراری تعادل بین اریبی برآوردگر ترکیبی و تغییرپذیری برآوردگر مستقیم در ناحیه‌های کوچک، استفاده از میانگین موزون این دو برآوردگر است. برآوردگر حاصل را برآوردگر مرکب گوئیم (راو، ۲۰۰۳). شکل کلی این برآوردگر به صورت

$$(۴) \quad \hat{p}_{i,C} = w_i \hat{p}_{i,D} + (1 - w_i) \hat{p}_{i,S_Y}$$

است، که $\hat{p}_{i,D}$ برآوردگر مستقیم مشخصه‌ی p در ناحیه‌ی کوچک i ام، و $\hat{p}_{i,S_Y}$ برآوردگر ترکیبی مشخصه‌ی p در ناحیه‌ی کوچک i ام است، که در این مقاله، مشخصه‌ی مورد نظر، نرخ بی‌کاری است. همچنین $0 \leq w_i \leq 1$ ، و مقدار بهینه برای آن از رابطه‌ی $w_i^* = \frac{1}{1+F_i}$ به دست می‌آید، که

$$(۵) \quad F_i = \frac{\text{MSE}(\hat{p}_{i,D})}{\text{MSE}(\hat{p}_{i,S_Y})};$$

یعنی، مقدار بهینه‌ی وزن w_i^* فقط به نسبت MSEها بستگی دارد. حال اگر برآورد مستقیم در ناحیه‌ی کوچک i ام نااریب باشد، رابطه‌ی (۵) به شکل $F_i = \text{var}(\hat{p}_{i,D}) / \text{MSE}(\hat{p}_{i,S_Y})$ خواهد بود.

جدول ۱. گروه‌بندی استان‌ها برای برآوردگرهای مرکب و جیمز-استاین

گروه	استان	گروه	استان
۱	آذربایجان شرقی	۳	آذربایجان غربی
	تهران		گیلان
	خراسان		فارس
	زنجان		مازندران
	سمنان		
	یزد	۴	ایلام
			خوزستان
۲	اصفهان		کرمانشاه
	بوشهر		کهگیلویه و بویراحمد
	چهارمحال و بختیاری		لرستان
	سیستان و بلوچستان		
	کردستان		
	کرمان		
	مرکزی		
	هرمزگان		
	همدان		

با توجه به این‌که در عمل، محاسبه‌ی برآورد $MSE(\hat{p}_{i,Sy})$ کار ساده‌ای نیست، با استفاده از برآورد تقریبی آن می‌توان رابطه‌ای به شکل

$$(۶) \quad \hat{w}_i^* = \frac{\widehat{MSE}(\hat{p}_{i,Sy})}{(\hat{p}_{i,Sy} - \hat{p}_{i,D})^2}$$

برای محاسبه‌ی w_i^* به دست آورد، که

$$(۷) \quad \widehat{MSE}(\hat{p}_{i,Sy}) \approx (\hat{p}_{i,Sy} - \hat{p}_{i,D})^2 - \text{var}(\hat{p}_{i,D}).$$

برآوردگر $\hat{w}_i^*$ دو مشکل اساسی دارد: اول این‌که بسیار ناپایدار و دارای تغییرات زیادی است، و دیگر این‌که برای برآورد $\widehat{MSE}(\hat{p}_{i,Sy})$ ممکن است در برخی از ناحیه‌های کوچک، مقادیر منفی به دست آید، که این مقادیر، موجب منفی شدن $\hat{w}_i^*$ می‌شود که معنایی ندارد. برای غلبه بر این دو مشکل، یک روش مناسب، میانگین‌گیری از $\hat{w}_i^*$ ‌ها روی چند متغیر یا چند ناحیه‌ی کوچک مشابه یا هر دو است.

در این مقاله پس از گروه‌بندی ناحیه‌های کوچک در چهار گروه مشابه از نظر نرخ بی‌کاری (جدول ۱)، میانگین $\hat{w}_i^*$ ‌های به دست آمده در هر گروه را برای ناحیه‌های کوچک آن گروه به کار برده‌ایم. شایان ذکر

است که در برخی از این گروه‌ها مقدار $\hat{w}_i^*$ به دلیل منفی شدن $\widehat{\text{MSE}}(\hat{p}_{i,\text{Sy}})$ منفی شده است؛ لذا مقدار گم‌شده تلقی شده و در محاسبه‌ی میانگین منظور نشده است.

حالت خاصی از برآوردگرهای مرکب، استفاده از یک وزن مشترک به جای $\hat{w}_i^*$ است. برآوردگر حاصل را برآوردگر جیمز-استاین گویند. به این ترتیب، رابطه‌ی (۴) به شکل

$$(۸) \quad \hat{p}_{i,\text{JS}} = w\hat{p}_{i,\text{D}} + (1-w)\hat{p}_{i,\text{Sy}}$$

در خواهد آمد، که وزن مشترک w با استفاده از رابطه‌ی

$$w = \frac{\sum_i \text{MSE}(\hat{p}_{i,\text{Sy}})}{\sum_i \{\text{MSE}(\hat{p}_{i,\text{D}}) + \text{MSE}(\hat{p}_{i,\text{Sy}})\}}$$

محاسبه می‌شود. ضمن این‌که برای محاسبه‌ی مقدار w در رابطه‌ی (۸) نیز باید از مقدار (۷) به جای $\text{MSE}(\hat{p}_{i,\text{Sy}})$ استفاده شود.

۲/۳ برآوردگر ترکیبی تعدیل‌شده‌ی رگرسیونی

این برآوردگر با ترکیب کردن متغیرهای کمکی معلوم در سطح هر ناحیه‌ی کوچک با برآوردگر ترکیبی آن ناحیه، تغییرپذیری ناحیه‌ای را در محاسبه‌ی برآورد به حساب می‌آورد. متغیرهای کمکی z_i برای مدل‌بندی اریبی نسبی مربوط به برآوردگر ترکیبی $\hat{p}_{i,\text{Sy}}$ در ناحیه‌ی کوچک i ام، یعنی $B_i = (p_i - \hat{p}_{i,\text{Sy}})/\hat{p}_{i,\text{Sy}}$ به صورت زیر به کار می‌روند (p_i مقدار واقعی نسبت مورد نظر در ناحیه‌ی کوچک i ام است).

$$(۹) \quad B_i = \gamma_0 + \gamma^T z_i + \varepsilon_i,$$

که در آن، γ_0 و γ به ترتیب عرض از مبدأ و بردار پارامترهای رگرسیونی اند، ε_i خطای تصادفی است، و z_i نیز بردار متغیرهای کمکی است. از آن‌جا که مقدار B_i قابل مشاهده نیست، ابتدا در ناحیه‌ی بزرگ‌تر a ، ($a = 1, \dots, A$)، مدل رگرسیونی را برای مقادیر برآورد شده‌ی $\hat{B}_a = (\hat{p}_a - \hat{p}_{a,\text{Sy}})/\hat{p}_{a,\text{Sy}}$ به روش کم‌ترین توان‌های دوم خطا برازش داده، پارامترهای مدل را محاسبه می‌کنیم ($\hat{p}_a$ برآورد نسبت در ناحیه‌ی a است و $\hat{p}_{a,\text{Sy}}$ برآورد ترکیبی نسبت در ناحیه‌ی a). در این صورت، $\hat{B}_a$ یک برآوردگر معتبر از $B_a = (p_a - \hat{p}_{a,\text{Sy}})/\hat{p}_{a,\text{Sy}}$ است. مقادیر برآورد شده‌ی عرض از مبدأ و بردار پارامترهای رگرسیونی را به ترتیب با $\hat{\gamma}_0$ و $\hat{\gamma}$ نشان می‌دهیم. به این ترتیب، B_i را با $\hat{\gamma}_0 + \hat{\gamma}^T z_i$ برآورد می‌کنیم، که منجر به برآورد ترکیبی تعدیل‌شده‌ی رگرسیونی زیر برای p_i می‌شود.

$$(۱۰) \quad \hat{p}_{i,\text{Sy,Ad}} = \hat{p}_{i,\text{Sy}}(1 + \hat{\gamma}_0 + \hat{\gamma}^T z_i).$$

جدول ۲. گروه‌بندی استان‌ها برای برآوردگر ترکیبی تعدیل‌شده‌ی رگرسیونی

گروه	استان	گروه	استان
۱	تهران	۵	سیستان و بلوچستان
	سمنان		کردستان
	یزد		همدان
۲	آذربایجان شرقی	۶	آذربایجان غربی
	خراسان		فارس
	زنجان		مازندران
۳	اصفهان	۷	ایلام
	بوشهر		کهگیلویه و بویراحمد
	مرکزی		گیلان
۴	چهارمحال و بختیاری	۸	خوزستان
	کرمان		کرمانشاه
	هرمزگان		لرستان

در این مقاله، با توجه به این‌که ناحیه‌های کوچک، استان‌های کشور هستند، کل استان‌ها به هشت گروه تقسیم شده (جدول ۲) و برای هر گروه، پس از محاسبه‌ی برآورد ترکیبی و برآورد مستقیم نرخ بی‌کاری، پارامترهای مدل با استفاده از متغیرهای کمکی در هر گروه محاسبه شده‌اند. پس از محاسبه‌ی این پارامترها، با استفاده از رابطه‌ی (۱۰) برآورد نرخ بی‌کاری برای هر استان به دست می‌آید.

۳ نکات عملی و محاسباتی

در این مقاله با استفاده از اطلاعات موجود در مورد نسبت جمعیت بیکار در کل کشور و به‌کارگیری روش سیستماتیک خطی، نمونه‌ای به اندازه‌ی ۱۵,۴۰۰ نفر در سطح اطمینان ۹۵٪ و حد اکثر خطای نسبی ۵٪ از کل جمعیت فعال کشور (جامعه‌ی تمام افراد ده‌ساله و بیش‌تر که طبق تعریف کار، شاغل یا بیکارند) در فایل سرشماری عمومی نفوس و مسکن ۱۳۷۵ انتخاب شده است.

به‌منظور فراهم کردن امکان مقایسه‌ی روش‌های برآورد برای ناحیه‌های کوچک با استفاده از معیارهای کیفیت که در بخش بعد معرفی خواهند شد، انتخاب نمونه‌ی مذکور ۱۰۰٪ مرتبه تکرار شده است. به این ترتیب، اندازه‌ی نمونه‌ای که به هر ناحیه‌ی کوچک (هر استان) اختصاص یافته، یک متغیر تصادفی است که

در هر تکرار می‌تواند متفاوت با تکرارهای دیگر باشد. از آن‌جا که در سرشماری عمومی نفوس و مسکن ۱۳۷۵، وضعیت اشتغال افراد، مورد پرسش قرار گرفته است، با استفاده از اطلاعات آن سال، نرخ بی‌کاری واقعی در سطح استان معلوم است.

با توجه به این‌که برآوردها برای ناحیه‌های کوچک از اطلاعات کمی بهره می‌گیرند و در کشور ما فاصله‌ی زمانی نسبتاً طولانی بین دو سرشماری نفوس و مسکن موجب کهنه شدن اطلاعات کمی می‌شود، به‌منظور سنجش تأثیر کهنگی اطلاعات کمی بر عملکرد برآوردها برای ناحیه‌ی کوچک، از اطلاعات کمی مربوط به سال ۱۳۶۵ استفاده شده است. اطلاعات کمی لازم برای انجام این بررسی در سطح استان‌ها و همچنین هر یک از پساتبقة‌ها، از پایگاه اطلاعات سرشماری عمومی نفوس و مسکن ۱۳۶۵ استخراج شده است. البته به‌دلیل تغییراتی که در فاصله‌ی زمانی ده‌ساله‌ی ۱۳۶۵ تا ۱۳۷۵ در استان‌ها ایجاد شده است، اطلاعات مربوط به استان‌های جدید در سال ۱۳۷۵ به استان‌های سابق خود منتسب شده و نتایج نیز بر اساس تقسیمات کشوری سال ۱۳۶۵ به دست آمده است. اطلاعات کمی مورد استفاده در این بررسی، دو گروه از اطلاعات را شامل می‌شود:

۱. اطلاعاتی که برای محاسبه‌ی برآوردهای ترکیبی به کار می‌رود و عبارت است از تعداد جمعیت فعال/بیکار در ناحیه‌های کوچک و پساتبقة‌ها؛

۲. اطلاعات مربوط به متغیرهای کمی، که برای محاسبه‌ی برآورد ترکیبی تعدیل‌شده‌ی رگرسیونی به کار می‌رود. متغیرهای اخیر از بین متغیرهایی که انتظار می‌رود رابطه‌ی قوی‌تری با نرخ بی‌کاری داشته باشند انتخاب شده‌اند.

برای محاسبه‌ی برآوردهای ترکیبی، لازم است جامعه را بر اساس یک مشخصه‌ی جمعیتی مرتبط با نرخ بی‌کاری به تعدادی زیرگروه (پساتبقة) تقسیم کنیم به‌طوری‌که اطلاعات کمی تعداد را در هر یک از این پساتبقة‌ها در سطح کل کشور و همچنین در سطح رده‌های متقابل پساتبقة‌ها و ناحیه‌های کوچک در اختیار داشته باشیم. به این ترتیب، جمعیت فعال کشور را بر اساس متغیر سن به شش گروه (۱۴-۱۰، ۱۹-۱۵، ۲۴-۲۰، ۲۹-۲۵، ۳۹-۳۰، و ۵۰+ ساله و بیش‌تر) تقسیم کرده‌ایم.

۴ مقایسه‌ی روش‌ها

همان‌طور که اشاره شد، به‌منظور مقایسه‌ی برآوردهای معرفی‌شده برای ناحیه‌های کوچک و فراهم آوردن امکان محاسبه‌ی معیارهای کیفیت برای این برآوردها، ابتدا تعداد ۱۰۰۰ نمونه‌ی تصادفی ساده با اندازه‌ی

۱۵,۴۰۰ نفر از فایل اطلاعات سرشماری عمومی نفوس و مسکن ۱۳۷۵ استخراج و سپس نرخ بی‌کاری در استان‌های کشور محاسبه شد. از آن‌جا که برای محاسبه‌ی برآوردها، مراجعه به نمونه‌ها به‌منظور گردآوری داده‌ها مورد نیاز نبود و پراکندگی نمونه‌ها هزینه‌ای در بر نداشت، برای به دست آوردن نمونه‌ای که معرف جامعه باشد، از روش نمونه‌گیری تصادفی ساده استفاده شد. با توجه به در دست بودن چارچوبی از واحدهای نمونه، امکان استفاده از این روش فراهم بود. محاسبه‌ی برآوردها در سطوح ناحیه‌های کوچک (استان‌ها) با استفاده از روش‌های برآورد برای ناحیه‌های کوچک، که در بخش‌های قبل توضیح داده شده است، انجام گرفت. برای محاسبه‌ی این برآوردها اطلاعات کمکی از قبیل تعداد در برخی از زیرگروه‌های جمعیتی یا مقدار برخی از پارامترهای جامعه در ناحیه‌های کوچک مورد نیاز بود، که این اطلاعات نیز از فایل سرشماری عمومی نفوس و مسکن ۱۳۶۵ استخراج شد و مورد استفاده قرار گرفت. استفاده از این اطلاعات، امکان بررسی تأثیر کهنگی اطلاعات کمکی بر برآوردها برای ناحیه‌های کوچک مورد بحث در این مقاله را فراهم ساخت. با توجه به معلوم بودن نرخ واقعی بی‌کاری در استان‌های کشور از فایل چارچوب و تعداد زیاد تکرارها، امکان محاسبه یا برآورد معیارهای مهم کیفیت برای برآوردها امکان‌پذیر است. از آن‌جا که اندازه‌ی نمونه‌ی مورد نیاز در این بررسی، در سطح کل کشور بهینه شده و تعداد نمونه‌ی اختصاص یافته به هر ناحیه‌ی کوچک (استان) یک متغیر تصادفی است، تأثیر اندازه‌ی نمونه بر مشخصه‌های کیفیت هر برآوردها پس از محاسبه‌ی مقدار مورد انتظار اندازه‌ی نمونه در ناحیه‌های کوچک، قابل بررسی است. همه‌ی محاسبات در این تحقیق با استفاده از برنامه‌نویسی به‌وسیله‌ی نرم‌افزار SAS انجام شده است. با توجه به تعداد نسبتاً زیاد برآوردهای محاسبه‌شده و همچنین ماهیت تحقیق به‌لحاظ استفاده از نمونه‌های تکراری، برای انجام دادن مقایسه‌ها و انتخاب برآوردهای کارا، از نمودارهای مختلف استفاده شده است.

۴/۱ معیارهای کیفیت

به‌منظور مقایسه‌ی کیفیت روش‌های مختلف برآورد برای ناحیه‌های کوچک، از چهار معیار استفاده شده است. باید توجه داشت که با فرض عدم وقوع خطاهای غیر نمونه‌گیری مانند خطای بی‌پاسخی یا خطای اندازه‌گیری در سرشماری عمومی نفوس و مسکن ۱۳۷۵، مقدار واقعی پارامتر جامعه (نرخ بی‌کاری) در سال ۱۳۷۵ معلوم است و با توجه به نمونه‌های تکراری که از چارچوب حاصل از فایل این سرشماری استخراج شده است، این معیارها قابل محاسبه و برآوردند.

آ) اریبی که تفاوت بین میانگین توزیع نمونه‌ای یک برآورد و مقدار واقعی پارامتر نامعلوم جامعه است. اگر $\hat{p}_i$ برآورد نرخ بی‌کاری در ناحیه‌ی کوچک i ام باشد و p_i نیز مقدار واقعی نرخ بی‌کاری این ناحیه‌ی کوچک از سرشماری عمومی نفوس و مسکن ۱۳۷۵ باشد، برآورد اریبی

نرخ بی‌کاری برای ناحیه‌ی کوچک i ام عبارت است از

$$(۱۱) \quad \widehat{\text{Bias}}(\hat{p}_i) = E(\hat{p}_i) - p_i = \frac{\sum_{j=1}^{1000} \hat{p}_{ij}}{1000} - p_i,$$

که $\hat{p}_{ij}$ نشان‌دهنده‌ی برآورد نرخ بی‌کاری ناحیه‌ی کوچک i ام در j امین نمونه‌ی انتخاب‌شده است.

ب) میانگین توان دوم خطا که میانگین توان دوم تفاوت‌های بین مقادیر برآورد شده و مقدار واقعی برای یک پارامتر نامعلوم جامعه در همه‌ی نمونه‌های ممکن از آن جامعه است و برآورد آن برای پارامتر مورد نظر عبارت است از

$$(۱۲) \quad \widehat{\text{MSE}}(\hat{p}_i) = E(\hat{p}_i - p_i)^2 = \sum_{j=1}^{1000} \frac{(\hat{p}_{ij} - p_i)^2}{1000}.$$

پ) متوسط خطای نسبی (ARE) و متوسط توان دوم خطا (ASE) که توسط گوش و راثو (۱۹۹۴) به منظور مقایسه‌ی عملکرد روش‌های برآورد برای ناحیه‌ی کوچک معرفی شده‌اند، عبارت‌اند از

$$(۱۳) \quad \text{ARE} = 100 \times \frac{1}{I} \sum_{i=1}^I \left(\frac{\hat{p}_i - p_i}{p_i} \right),$$

و

$$(۱۴) \quad \text{ASE} = \frac{1}{I} \sum_{i=1}^I (\hat{p}_i - p_i)^2,$$

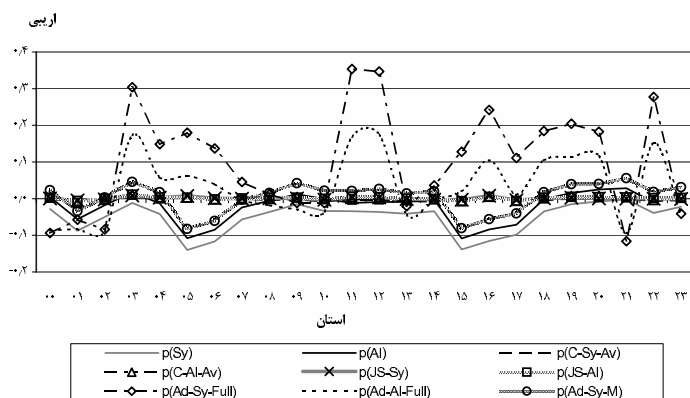
که I تعداد ناحیه‌های کوچک، $\hat{p}_i$ برآورد نرخ بی‌کاری در ناحیه‌ی کوچک i ام، و p_i مقدار واقعی نرخ بی‌کاری در ناحیه‌ی کوچک i ام است. شایان ذکر است که در این بررسی برای هر برآوردگر، به تعداد تکرارها (هزار تکرار) ARE و ASE به دست می‌آید. به این ترتیب، برای مقایسه‌ی برآوردگرها از نظر این معیارها، از نمودار شمایی استفاده می‌شود.

۴/۲ بحث و نتیجه‌گیری

نخست نمادهایی که در شکل‌ها مورد استفاده قرار گرفته‌اند معرفی می‌شوند.
 $p(\text{Sy})$: برآوردگر ترکیبی،

$p(Al)$: برآوردگر ترکیبی جایگزین،
 $p(C-Sy)$: برآوردگر مرکب با استفاده از برآوردهای ترکیبی و MSE واقعی،
 $p(C-Al)$: برآوردگر مرکب با استفاده از برآوردهای ترکیبی جایگزین و MSE واقعی،
 $p(C-Sy-Av)$: برآوردگر مرکب با استفاده از برآوردگر ترکیبی و میانگین وزن گروه‌ها،
 $p(C-Al-Av)$: برآوردگر مرکب با استفاده از برآوردگر ترکیبی جایگزین و میانگین وزن گروه‌ها،
 $p(JS-Sy)$: برآوردگر جیمز-استاین با استفاده از برآوردگر ترکیبی،
 $p(JS-Al)$: برآوردگر جیمز-استاین با استفاده از برآوردگر ترکیبی جایگزین،
 $p(Ad-Sy-Full)$: برآوردگر ترکیبی تعدیل‌شده‌ی رگرسیونی با استفاده از برآوردگر ترکیبی و مجموعه‌ی کامل متغیرهای کمکی موجود،
 $p(Ad-Al-Full)$: برآوردگر ترکیبی تعدیل‌شده‌ی رگرسیونی با استفاده از برآوردگر ترکیبی جایگزین و مجموعه‌ی کامل متغیرهای کمکی موجود،
 $p(Ad-Sy-M)$: برآوردگر ترکیبی تعدیل‌شده‌ی رگرسیونی با استفاده از برآوردگر ترکیبی و میانه‌ی سنی جمعیت بیکار به‌عنوان مهم‌ترین متغیر کمکی.

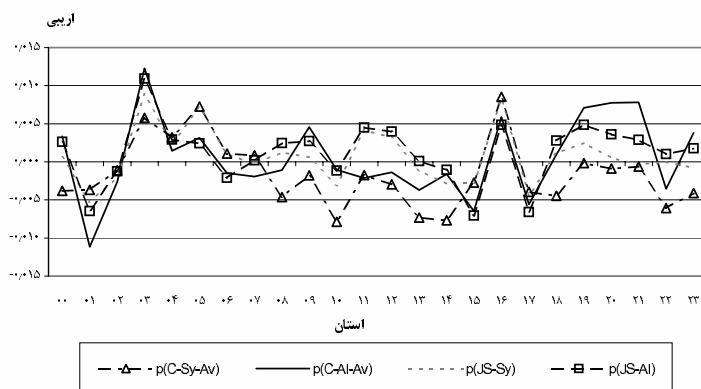
به‌منظور مقایسه‌ی روش‌های برآورد برای ناحیه‌های کوچک مورد بحث در بخش ۲، از شکل‌های متعددی استفاده شده است. شکل ۱ اریبی همه‌ی برآوردها برای ناحیه‌های کوچک در استان‌های مختلف را در کنار یکدیگر نشان می‌دهد. همان‌طور که در این شکل مشاهده می‌شود، برآوردگر ترکیبی $p(Sy)$ در بین این برآوردها دارای اریبی بسیار زیادی است و همواره موجب کم‌برآوردی می‌شود. با توجه به شکل (۱) و حذف برآوردهای دارای اریبی زیاد (شکل ۲) مشخص می‌شود برآوردهای



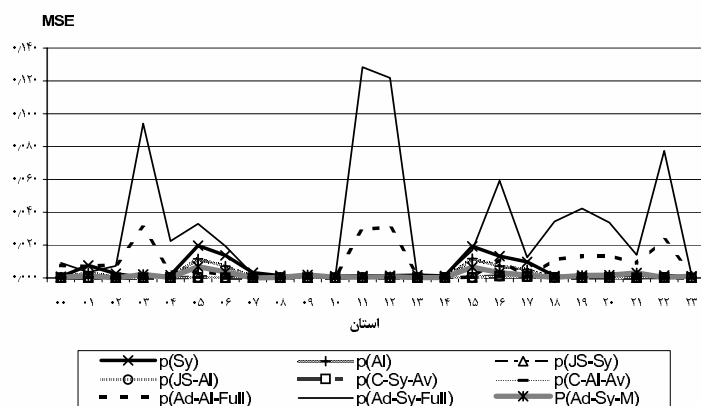
شکل ۱. اریبی روش‌های برآورد برای ناحیه‌های کوچک

مرکب $(p(C-Sy-Av))$ و $(p(C-Al-Av))$ و جیمز-استاین $(p(JS-Sy))$ و $(p(JS-Al))$ نسبت به سایر برآوردگرها اریبی کمتری دارند.

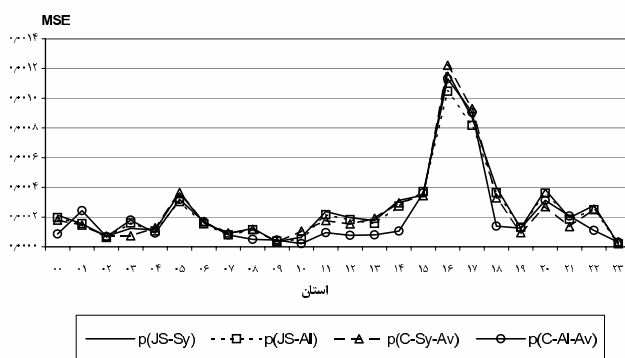
با توجه به شکل ۳ که MSE مربوط به برآوردگرها برای ناحیه‌ی کوچک در استان‌های مختلف را در کنار یکدیگر نشان می‌دهد، چهار برآوردگر $(p(JS-Al))$ ، $(p(JS-Sy))$ ، $(p(C-Al-Av))$ و $(p(C-Sy-Av))$ نسبت به برآوردگرهای دیگر از MSE کمتری برخوردارند. با حذف سایر برآوردگرها، می‌توان MSE این چهار برآوردگر را در شکل ۴ در کنار یکدیگر مشاهده کرد.



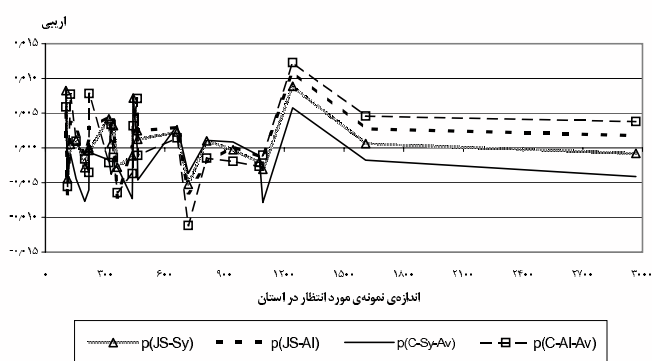
شکل ۲. اریبی چهار برآوردگر برگزیده برای ناحیه‌های کوچک



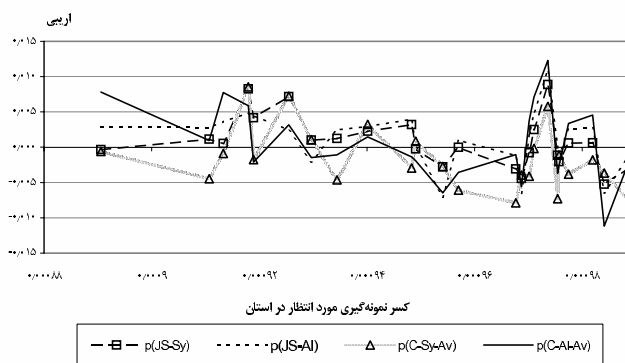
شکل ۳. MSE هر یک از روش‌های برآورد برای ناحیه‌های کوچک



شکل ۴. MSE برآوردگرهای برگزیده برای ناحیه‌های کوچک



شکل ۵. اریبی چهار برآوردگر برگزیده برای ناحیه‌های کوچک در مقابل اندازه‌ی نمونه‌ی مورد انتظار در استان



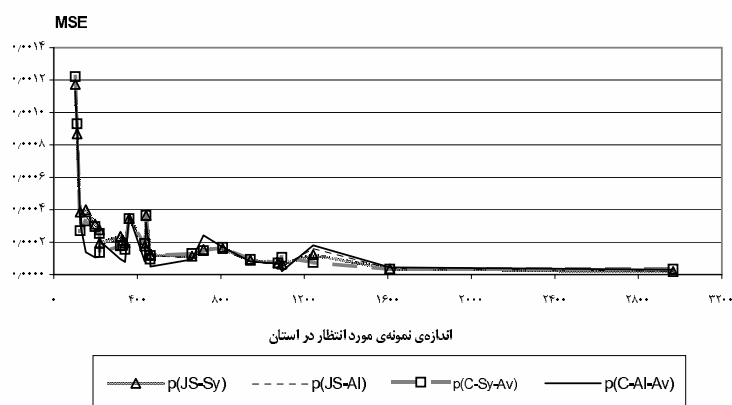
شکل ۶. اریبی چهار برآوردگر برگزیده برای ناحیه‌های کوچک در مقابل کسر نمونه‌گیری مورد انتظار در استان

جدول ۳. متوسط اندازه‌ی نمونه و کسر نمونه‌گیری به تفکیک استان

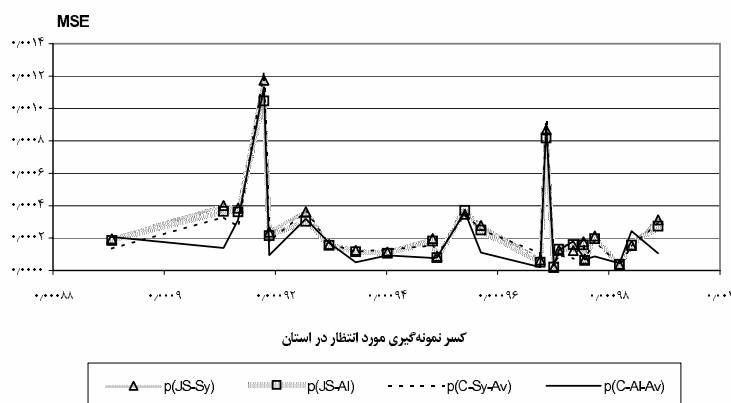
کد استان	نام استان	متوسط اندازه‌ی نمونه	متوسط کسر نمونه‌گیری
۰۰	مرکزی	۳۳۰	۰/۰۰۰۰۹۷۷
۰۱	گیلان	۷۱۷	۰/۰۰۰۰۹۸۴
۰۲	مازندران	۱۰۷۴	۰/۰۰۰۰۹۷۶
۰۳	آذربایجان شرقی	۱۲۴۲	۰/۰۰۰۰۹۷۴
۰۴	آذربایجان غربی	۶۶۲	۰/۰۰۰۰۹۴۰
۰۵	کرمانشاه	۴۴۱	۰/۰۰۰۰۹۲۵
۰۶	خوزستان	۸۱۰	۰/۰۰۰۰۹۳۰
۰۷	فارس	۹۴۲	۰/۰۰۰۰۹۴۹
۰۸	کرمان	۴۶۳	۰/۰۰۰۰۹۳۴
۰۹	خراسان	۱۶۰۹	۰/۰۰۰۰۹۸۲
۱۰	اصفهان	۱۰۹۲	۰/۰۰۰۰۹۶۸
۱۱	سیستان و بلوچستان	۳۱۹	۰/۰۰۰۰۹۱۹
۱۲	کردستان	۳۴۱	۰/۰۰۰۰۹۴۸
۱۳	همدان	۴۳۶	۰/۰۰۰۰۹۷۵
۱۴	چهارمحال و بختیاری	۱۹۸	۰/۰۰۰۰۹۸۹
۱۵	لرستان	۳۶۱	۰/۰۰۰۰۹۵۴
۱۶	ایلام	۱۰۳	۰/۰۰۰۰۹۱۸
۱۷	کهگیلویه و بویراحمد	۱۱۱	۰/۰۰۰۰۹۶۹
۱۸	بوشهر	۱۵۴	۰/۰۰۰۰۹۱۱
۱۹	زنجان	۴۶۱	۰/۰۰۰۰۹۷۱
۲۰	سمنان	۱۲۶	۰/۰۰۰۰۹۱۳
۲۱	یزد	۲۲۰	۰/۰۰۰۰۸۹۱
۲۲	هرمزگان	۲۱۸	۰/۰۰۰۰۹۵۷
۲۳	تهران	۲۹۶۷	۰/۰۰۰۰۹۷۰

همان‌طور که در شکل ۴ ملاحظه می‌شود، برای استان‌های ایلام و کهگیلویه و بویراحمد (استان‌های ۱۶ و ۱۷) افزایش زیادی در مقدار MSE در تمام روش‌ها رخ داده است. به نظر می‌رسد که این افزایش در مقدار MSE بی‌ارتباط با اندازه‌ی نمونه در این استان‌ها نباشد. جدول ۳ مقدار مورد انتظار اندازه‌ی نمونه و همچنین کسر نمونه‌گیری مربوط به هر استان را نشان می‌دهد. با توجه به این جدول، استان‌های مذکور دارای کم‌ترین مقدار مورد انتظار برای اندازه‌ی نمونه هستند. آریبی چهار روش برگزیده در مقابل اندازه‌ی نمونه و کسر نمونه‌گیری مورد انتظار، به‌ترتیب در شکل‌های ۵ و ۶ رسم شده است. با بررسی این شکل‌ها

ملاحظه می‌شود که اریبی برآوردگرها با افزایش اندازه‌ی نمونه و کسر نمونه‌گیری، به‌جز در دو مورد که اندازه‌ی نمونه‌ی بزرگی را به خود اختصاص داده‌اند، روند خاصی از خود نشان نمی‌دهد. شکل ۷ مقدار MSE چهار روش برگزیده را در مقابل اندازه‌ی نمونه‌ی مورد انتظار برای استان‌های مختلف نشان می‌دهد. با توجه به این شکل، این‌گونه استنباط می‌شود که در اغلب موارد با افزایش اندازه‌ی نمونه در یک ناحیه‌ی کوچک، میانگین توان دوم خطای برآوردگر ناحیه‌ی کوچک برای آن ناحیه کاهش می‌یابد. در این شکل، MSE برآوردگرها برای اندازه‌های مورد انتظار کم‌تر از ۴۵۰، ناپایدار است و مقادیر بسیار بزرگی را اختیار می‌کند؛



شکل ۷. MSE چهار برآوردگر برگزیده برای ناحیه‌های کوچک در مقابل اندازه‌ی نمونه‌ی مورد انتظار در استان



شکل ۸. MSE چهار برآوردگر برگزیده برای ناحیه‌های کوچک در مقابل کسر نمونه‌گیری مورد انتظار در استان

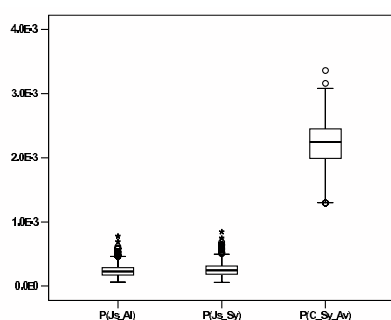
ولی برای مقادیر مورد انتظار بیش از 45° ، علاوه بر این‌که روندی تقریباً کاهشی دارد، به حالتی تقریباً پایدار نیز رسیده است. با نادیده گرفتن برآوردگر $p(C-AI-AV)$ ، مشاهده می‌شود که برای مقادیر مورد انتظار بیش از 45° ، همواره سه برآوردگر برگزیده دیگر، مقدار MSE کم‌تر از 17×10^{-6} دارند. البته در یک حالت ایده‌آل برای اندازه‌ی نمونه‌ی بیش از 80° همواره برای هر چهار روش، مقدار MSE کم‌تر از 2×10^{-6} به دست آمده است.

شکل ۸ نشان می‌دهد که مقادیر MSE نسبت به افزایش کسر نمونه‌گیری مورد انتظار برای استان‌ها، رفتار منظمی از خود نشان نمی‌دهند. به این ترتیب، فارغ از کسر نمونه‌گیری مورد انتظار برای یک ناحیه‌ی کوچک، اندازه‌ی نمونه‌ی آن ناحیه‌ی کوچک در دست‌یابی به برآوردهایی با میانگین توان دوم خطای مطلوب، حائز اهمیت است.

با بررسی مقادیر به دست آمده برای معیار ARE، برآوردهای مرکب $(p(C-Sy-AV))$ و جیمز-استاین $(p(JS-AI))$ و $(p(JS-Sy))$ برآوردهای برگزیده‌ای هستند که در بین سایر برآوردها از مقادیر ARE کم‌تری برخوردارند. شکل ۹ یک نمودار شمایی است که مقادیر ARE برای این سه برآوردگر را در کنار یکدیگر نشان می‌دهد. در این شکل، مشاهده می‌شود که برآوردهای جیمز-استاین نسبت به برآوردگر مرکب از مقادیر ARE کوچک‌تری برخوردارند؛ ضمن این‌که هر سه برآوردگر از نظر توزیع مقادیر ARE شکل یکسانی دارند. برای روش‌های جیمز-استاین، نقاط دورافتاده مشاهده می‌شود؛ اما این نقاط همواره از مینیمم مقدار ARE در روش مرکب کوچک‌ترند.

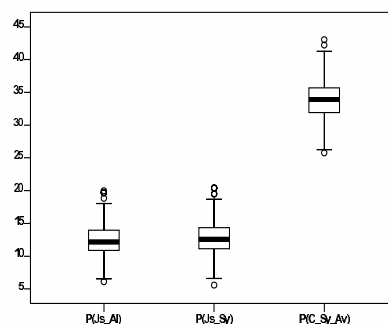
شکل ۱۰ نتایج مشابه با شکل ۹ را برای معیار ASE نشان می‌دهد.

برای محاسبه‌ی برآوردگر ترکیبی تعدیل‌شده‌ی رگرسیونی، متغیرهای کمکی‌ای که انتظار می‌رود رابطه‌ی قوی‌تری با نرخ بی‌کاری داشته باشند انتخاب شده‌اند. به این منظور، متغیرهای میانگین سنی جمعیت



شکل ۱۰. مقادیر ASE در هر یک از تکرارهای

نمونه‌گیری، به تفکیک برآوردهای برگزیده

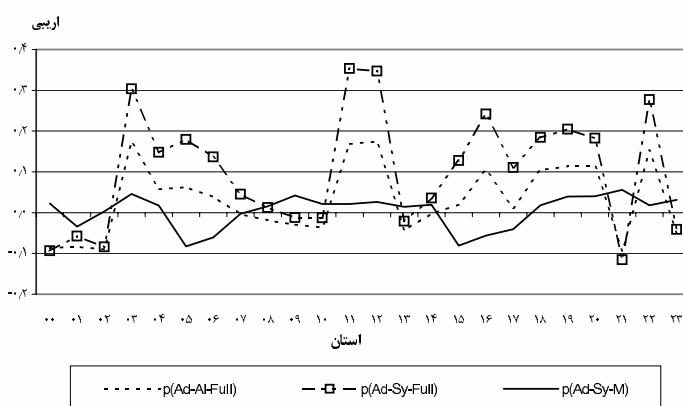


شکل ۹. مقادیر ARE در هر یک از تکرارهای

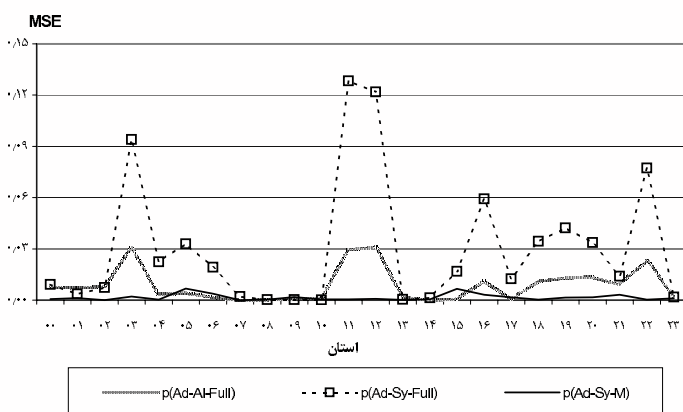
نمونه‌گیری، به تفکیک برآوردهای برگزیده

فعال، تعداد مردان در جمعیت فعال، تعداد افراد باسواد در جمعیت فعال، تعداد افراد شهری در جمعیت فعال، میانه‌ی سنی جمعیت فعال، و میانه‌ی سنی جمعیت بیکار، به تفکیک ناحیه‌های کوچک (استان‌ها) و همچنین گروه‌های بزرگ‌تر حاصل از این ناحیه‌های کوچک، از فایل سرشماری عمومی نفوس و مسکن ۱۳۶۵ استخراج شد.

ترکیب‌های متفاوتی از متغیرهای کمکی در مدل 10° به کار برده شد و نتیجه با استفاده از برآوردهای ترکیبی معمولی و ترکیبی جایگزین محاسبه شد. اریبی و MSE برآوردهای منتخب رگرسیونی را می‌توان در شکل‌های ۱۱ و ۱۲ مشاهده کرد. ضریب همبستگی متغیرهای کمکی در مقابل نرخ بی‌کاری



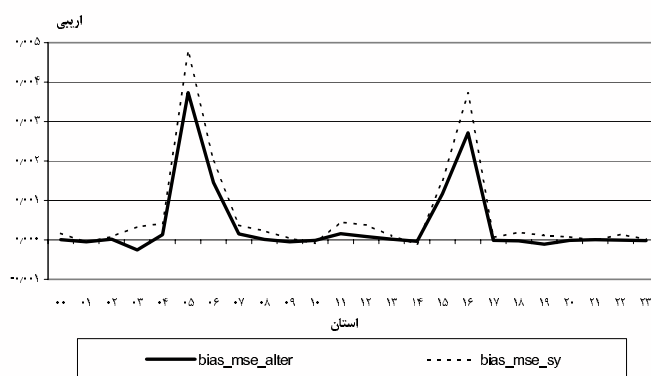
شکل ۱۱. اریبی روش‌های منتخب برآورد ترکیبی تعدیل‌شده‌ی رگرسیونی



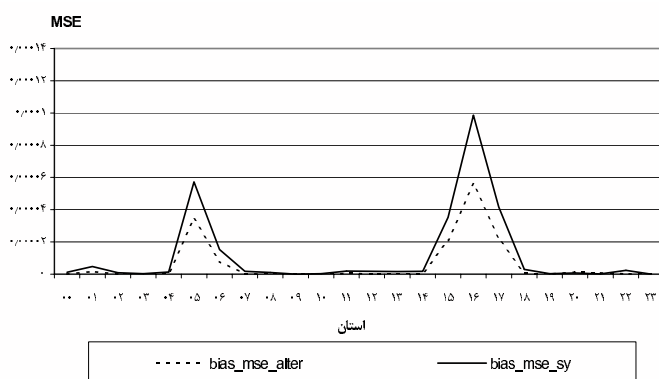
شکل ۱۲. MSE روش‌های منتخب برآورد ترکیبی تعدیل‌شده‌ی رگرسیونی

واقعی استان‌ها نشان می‌دهد که میانه‌ی سنی جمعیت بیکار، رابطه‌ی قوی‌تری نسبت به سایر متغیرهای کمکی با نرخ بی‌کاری دارد. در شکل‌ها نیز دیده می‌شود که برآوردگر متناظر، از نظر MSE و اریبی، نسبت به دیگر برآوردگرهای ترکیبی تعدیل‌شده‌ی رگرسیونی وضعیت بهتری دارد. اما به‌طور کلی، برآوردگرهای مرکب و جیمز-استاین نسبت به مجموعه‌ی برآوردگرهای ترکیبی رگرسیونی تعدیل‌شده، از نظر معیارهایی که در این تحقیق مورد بررسی قرار گرفته‌اند، رفتار بهتری را از خود نشان می‌دهند.

با توجه به این‌که در رابطه‌ی (۴) محاسبه‌ی وزن‌های w_i مستلزم معلوم بودن مقدار عملاً نامعلوم میانگین توان دوم خطاها برای برآوردگرهای ترکیبی است، این مقدار از رابطه‌ی (۷) برآورد می‌شود. برای آگاهی از کیفیت این برآورد، اریبی و میانگین توان دوم خطاها برای برآوردگر (۷) نیز برآورد شده است، که در شکل‌های

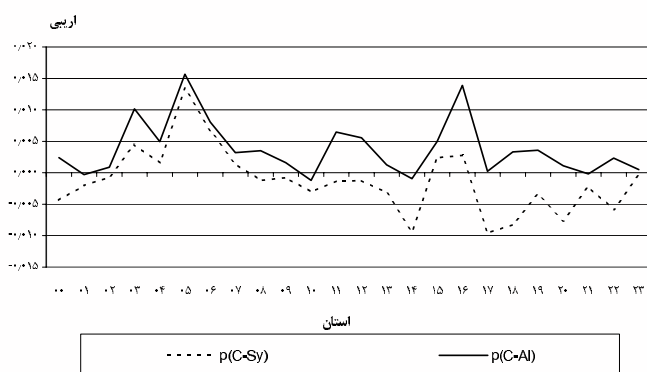


شکل ۱۳. اریبی برآورد MSE برآوردگرهای ترکیبی

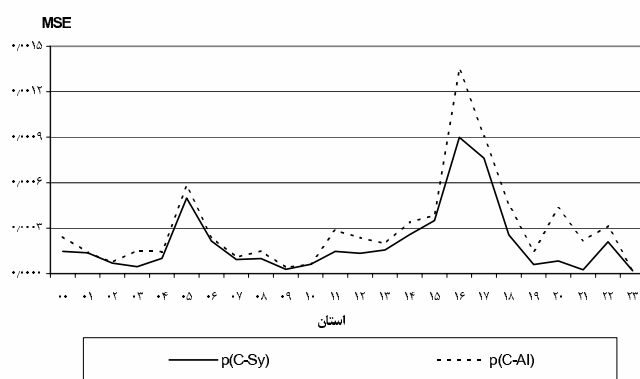


شکل ۱۴. میانگین توان دوم خطاها برای برآوردگر MSE برآوردگرهای ترکیبی

۱۳ و ۱۴ مشاهده می‌شوند. بررسی این شکل‌ها نشان می‌دهد که برآورد میانگین توان دوم خطاها برای برآوردگرهای ترکیبی از لحاظ اریبی و MSE وضعیت ناپایداری دارد، به‌طوری که برای برخی از ناحیه‌های کوچک، مقادیر اریبی و MSE بسیار بزرگی به دست آمده است. جالب توجه است که در هر دو شکل، بیش‌ترین مقادیر، متعلق به استان‌های کرمانشاه و ایلام (استان‌های ۵ و ۱۶) است، که برای این استان‌ها MSE برآوردگرهای برگزیده نیز مقادیر بزرگی اختیار کرده است. این شکل‌ها همچنین نشان می‌دهند که اریبی و MSE میانگین توان دوم خطاها برای برآوردگر ترکیبی جایگزین، مقادیر کم‌تری دارد. همان‌طور که قبلاً نیز اشاره شد، در این بررسی، امکان محاسبه‌ی MSE واقعی برآوردگرهای ترکیبی و ترکیبی جایگزین فراهم است. به این ترتیب می‌توان مقادیر برآورد مرکب را با به‌کارگیری برآوردهای ترکیبی و



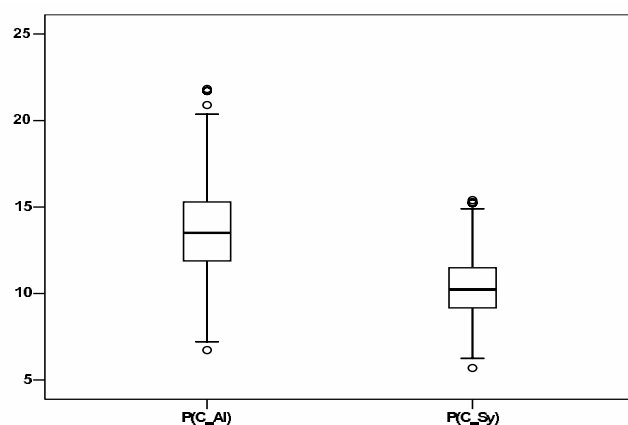
شکل ۱۵. اریبی برآوردگرهای مرکب با w_i^*



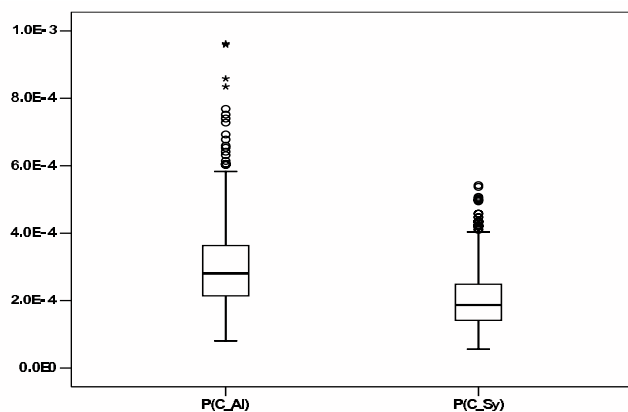
شکل ۱۶. MSE برآوردگرهای مرکب با w_i^*

ترکیبی جایگزین و وزن‌هایی که با استفاده از مقدار واقعی MSE این برآوردگرها به دست آمده‌اند، محاسبه کرد.

با مقایسه‌ی برآوردگرهای حاصل از این محاسبات، از نظر تمام معیارهای کیفیتی که در این تحقیق بررسی شده است، نتیجه می‌شود که برآوردگر مرکب حاصل از برآوردگر ترکیبی با استفاده از MSE واقعی $(p(C-Sy))$ ، نسبت به برآوردگرهای برگزیده برتری دارد (شکل‌های ۱۵ تا ۱۸)؛ اما از آن‌جا که در عمل، هیچ‌گاه مقدار MSE معلوم نیست و همان‌طور که در شکل‌های ۱۳ و ۱۴ دیده شد، برآورد آن نیز مقدار دقیقی نیست، امکان استفاده از برآوردگر مرکب با این شرایط مطلوب هیچ‌گاه وجود ندارد.

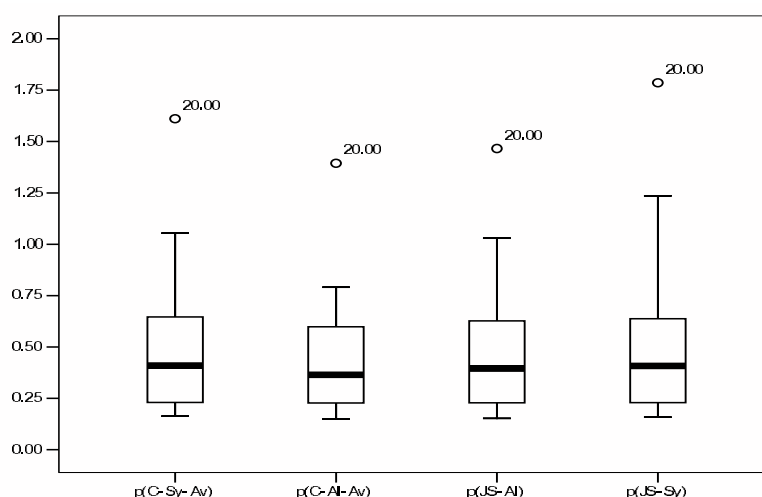


شکل ۱۷. ARE برآوردگرهای مرکب با w_i^*



شکل ۱۸. ASE برآوردگرهای مرکب با w_i^*

شکل ۱۹ نمودار شمایی حد اکثر خطای نسبی برآوردها را برای چهار روش برگزیده‌ی برآورد برای ناحیه‌های کوچک ($p(JS-Sy)$ ، $p(JS-Al)$ ، $p(C-Al-Av)$ و $p(C-Sy-Av)$) نشان می‌دهد. با مشاهده‌ی این نمودار، مشخص می‌شود که روش‌های برگزیده در فاصله‌ی بین چارک اول و سوم خطاها رفتار مشابهی دارند و تفاوت بین روش‌های مختلف، مربوط به نقاط کرانگین آن‌هاست. از سوی دیگر، یکسان بودن مشاهده‌ای که توسط نرم‌افزار در همه‌ی روش‌ها به‌عنوان داده‌ی دورافتاده تشخیص داده شده، نشان‌دهنده‌ی آن است که اگر برآورد مورد نظر (نرخ بی‌کاری) با استفاده از یکی از روش‌های برگزیده، دارای خطای نسبی زیادی باشد، این برآورد با استفاده از سه روش دیگر نیز دارای خطای نسبی زیادی خواهد بود. با وجود این، براساس داده‌های حاصل از نمونه‌ها، روش $p(C-Al-Av)$ در بین چهار روش برگزیده، از حداکثر خطای نسبی کم‌تری برخوردار است. به این ترتیب، با در نظر گرفتن یک مقدار مینیمم برای اندازه‌ی نمونه در هر ناحیه‌ی کوچک، به‌منظور محاسبه‌ی نرخ بی‌کاری در ناحیه‌های کوچک می‌توان با استفاده از روش‌های $p(C-Sy-Av)$ ، $p(C-Al-Av)$ ، $p(JS-Sy)$ و $p(JS-Al)$ برآوردهای قابل قبولی از لحاظ معیار MSE به دست آورد؛ ضمن این‌که این چهار برآوردگر نسبت به سایر روش‌ها مقدار آریبی کم‌تری دارند و در بین این چهار برآوردگر نیز از نظر ARE و ASE ، برآوردگرهای جیمز-استاین نسبت به برآوردگرهای مرکب، مقادیر کم‌تری را اختیار کرده‌اند.



شکل ۱۹. حد اکثر خطای نسبی برآوردها در ناحیه‌های کوچک برای چهار روش برگزیده

سپاس‌گزاری

در این بررسی، از نظرهای ارزشمند و راهنمایی‌های بی‌دریغ آقای محمدباقر سخاوت بهره‌گرفته شده است، که صمیمانه از ایشان سپاس‌گزاری می‌شود. همچنین از خانم مریم زنگنه به پاس همکاری‌های بی‌شائبه و اراکه‌ی نظرهای مفید، و از آقای علی‌رضا خوشگویان‌فرد که با توضیحات مفید خود در مورد سابقه‌ی بررسی‌های انجام‌گرفته، ما را یاری کردند، تشکر و قدردانی می‌شود.

مرجع‌ها

- خوشگویان‌فرد، علی‌رضا (۱۳۸۰). استفاده از برآوردهای ترکیبی و برخی از برآوردهای مبتنی بر مدل برای برآورد نسبت بی‌کاری در سطح ناحیه‌های کوچک (برخی از استان‌ها). پایان‌نامه‌ی کارشناسی ارشد، دانشگاه علامه طباطبائی، تهران.
- لهوی، پل اس؛ لمی‌شو، استتلی (۱۹۹۹). نمونه‌گیری: روش‌ها و کاربردها. گیتی مختاری امیرمجدی (مترجم)، ترجمه از نسخه‌ی انگلیسی، پژوهشکده‌ی آمار، تهران، ۱۳۸۱.
- مرکز آمار ایران (۱۳۷۱). سالنامه‌ی آماری کشور-۱۳۷۰. مرکز آمار ایران، تهران.
- مرکز آمار ایران (۱۳۷۹). سالنامه‌ی آماری کشور-۱۳۷۸. مرکز آمار ایران، تهران.
- Rao, J.N.K. (2003). *Small Area Estimation*. Wiley, New York.
- Estevao, V.M.; Särndal, C.E. (2004). Borrowing strength is not the best technique within a wide class of design-consistent domain estimators. *J. Official Stat.* **20**, 645-669.
- Costa, A.; Satorra, A.; Ventura, E. (2004). Improve both domain and total area estimation by composition. *SORT* **28**, 69-86.
- Ghosh, M.; Rao, J.N.K. (1994). Small area estimation: an appraisal. *Statist. Sci.* **9**, 55-93.

روشنگر علی اکبری صبا

مرکز آمار ایران،

خیابان دکتر فاطمی، نبش خیابان رهی معیری،

تهران، ایران.

پایان‌نگار: r_aliakbari@sci.org.ir

آرمان بیداربختنیا

مرکز آمار ایران،

خیابان دکتر فاطمی، نبش خیابان رهی معیری،

تهران، ایران.

پایان‌نگار: a_bidar@sci.org.ir