

برآوردهای کوچک ناحیه‌ای نرخ بیکاری

فائزه عباس‌زاده* و حمیدرضا نواب‌پور

دانشگاه علامه طباطبائی

چکیده: نرخ بیکاری یکی از شاخص‌های مهم برای نمایش وضعیت اقتصادی-اجتماعی جامعه است. برای تخصیص مؤثر بودجه‌ها، اجرای برنامه‌های اشتغال و هدایت سیاست‌گذاری‌ها نیاز به داشتن برآوردهای قابل قبول برای پارامترهای بیکاری و اشتغال در سطح کشور هستیم. اداره‌های آمار ملی با این مشکل مواجه هستند که در نمونه‌گیری هنگامی که اندازه‌ی نمونه‌ای در سطح کشور بهینه می‌شود، اندازه‌های نمونه‌ای به‌منظور برآورد در ناحیه‌های کوچک‌تر از کشور، برآوردهای مستقیم قابل قبولی به دست نمی‌دهد و افزایش اندازه‌ی نمونه‌ای در بسیاری از ناحیه‌ها نیز ممکن نیست. بنا بر این لازم است تا از فن‌های برآورد کوچک‌ناحیه‌ای برای ساختن برآوردهای قابل قبول بهره‌گرفت. با استفاده از رویکرد مدل‌مبنا که در آن متغیرهای پاسخ از طریق یک مدل با متغیرهای کمکی ارتباط برقرار می‌کنند می‌توان دقت برآوردهای کوچک‌ناحیه‌ای را با وام‌گرفتن قدرت از این متغیرها بهبود بخشید. در این مقاله دو مدل رگرسیون لوژستیکی و مدل آمیخته‌ی لوجیتی چندجمله‌ای با اثرهای تصادفی برای برآورد نسبت بیکاری در استان‌ها معرفی می‌شوند. سپس در یک مطالعه‌ی شبیه‌سازی با استفاده از داده‌های طرح نیروی کار مرکز آمار ایران در سال ۱۳۸۹ دقت برآوردهای کوچک‌ناحیه‌ای مدل‌مبنا و طرح‌مبنا با یکدیگر مقایسه می‌شوند.

واژگان کلیدی: برآوردهای مستقیم؛ برآوردهای کوچک‌ناحیه‌ای؛ مدل رگرسیون لوژستیکی؛ مدل آمیخته‌ی لوجیتی چندجمله‌ای؛ آمارگیری نیروی کار.

* نویسنده‌ی عهده‌دار مکاتبات

دریافت: ۱۳۹۲/۱۲/۱۲، پذیرش: ۱۳۹۳/۹/۳.

۱- مقدمه

برای به دست آوردن آماره‌هایی با دقت مورد نظر، نیازمند آمارگیری از جامعه‌ی هدف هستیم. در بسیاری از کشورها برای به دست آوردن اطلاعات مورد نیاز، سرشماری هر ده سال یکبار انجام می‌گیرد. با توجه به رشد روزافزون جمعیت‌ها و تغییر در ویژگی‌های جمعیت‌شناختی آن‌ها در طول زمان، و روزآمد نبودن اطلاعات آماری در فاصله‌ی زمانی اجرای دو سرشماری، معمولاً اطلاعات رضایت‌بخشی برای سطح‌های جغرافیایی یا موضوعی کوچک وجود ندارد. همین مسئله باعث گسترش آمارگیری نمونه‌ای شده است. آمارگیری نمونه‌ای معمولاً برای ساختن برآوردهایی برای جامعه‌ی هدف و زیرگروه‌های اصلی آن طراحی می‌شوند. برآوردهای آمارگیرهای استاندارد برای گروه‌های اصلی، برآوردهای مستقیم یا طرح‌مبنا نامیده می‌شوند.

در نمونه‌گیری ممکن است به ناحیه‌هایی برخورد کنیم که به دلیل اندازه‌ی نمونه‌ای کم یا حتی صفر برآوردهای مستقیم مبتنی بر آن‌ها دقت کافی نداشته باشند. در این حالت نیازمند توسعه‌ی روش‌ها و فن‌های برآورد مبتنی بر اندازه‌ی نمونه‌ای کوچک هستیم. یک ناحیه «کوچک» نامیده می‌شود اگر اندازه‌ی نمونه‌ای برای برآوردهای مستقیم پارامترهای ناحیه به اندازه‌ی کافی بزرگ نباشد که دقت لازم را فراهم کند.

وجود نمونه‌ای جامع، شامل تمام ناحیه‌ها به طوری که اندازه‌ی نمونه‌ای مربوط به هر ناحیه برای پشتیبانی برآوردهای مستقیم به اندازه‌ی کافی بزرگ باشد، تقریباً غیر ممکن است. در نمونه‌گیری‌های کشوری ممکن است اندازه‌های نمونه‌های قرارگرفته در استان‌ها کوچک باشند و بنا بر این برآوردهای مستقیم مبتنی بر آن‌ها ناکارا هستند. در این حالت برای افزایش کارایی برآوردها از روش‌های برآورد کوچک‌ناحیه‌ای استفاده می‌شود. برای استفاده از فن برآورد کوچک‌ناحیه‌ای نیاز به داشتن اطلاعات کمکی موثق است. بنا بر این متغیرهای کمکی یا متغیرهای پیش‌گو نقش بسیار مهمی را در برآوردهای کوچک‌ناحیه‌ای ایفا می‌کنند. موفقیت روش‌های مدل‌مبنا به موجود بودن داده‌های کمکی مناسب بستگی دارد.

روش‌های برآورد کوچک‌ناحیه‌ای از طریق یک مدل آماری، رابطه‌ای را بین متغیر پاسخ و اطلاعات کمکی ایجاد می‌کنند که منجر به ساختن برآوردهای کوچک‌ناحیه‌ای مبتنی بر مدل یا برآوردها نامستقیم می‌شود. یاری گرفتن از اطلاعات کمکی، می‌تواند برآوردهایی

دقیق‌تر از براوردهای مستقیم سنتی (که تنها از داده‌های نمونه‌ای استفاده می‌کنند) حاصل کند.

در آمارگیری نیروی کار اگر وضع فعالیت فرد (شاغل یا بیکار) به صورت یک متغیر دو حالتی ($Y = 1$) در نظر گرفته شود به گونه‌ای که متغیر کمکی‌ای مثل X وجود داشته باشد که بر احتمال وقوع حالت ۱ آن تأثیرگذار باشد یعنی $P(Y = 1) = P(X)$ ؛ آن‌گاه دیدگاه‌های مختلفی برای تعیین ارتباط بین X و Y وجود دارد. یکی از این دیدگاه‌ها استفاده از تابع رگرسیون لوژیستیک است که در آن تابع لوجیتی $\log(p(x)/(1-p(x))) = \text{logit}(p(x))$ با یک تابع خطی از متغیرهای کمکی و اثرهای ثابت مرتبط است.

با استفاده از برآزش دو تابع مجزای رگرسیون لوژیستیک می‌توان براوردهایی را برای نسبت‌های بیکاری و اشتغال و همچنین نرخ بیکاری در استان‌ها به‌عنوان کوچک‌ناحیه‌ها به دست آورد. اما با وجود همبستگی بین مجموع شاغلان و بیکاران در کوچک‌ناحیه‌ها و همچنین وجود متغیر پاسخ سه حالتی وضع فعالیت فرد (شاغل و بیکار و غیر فعال اقتصادی بودن) مدل جدیدی با عنوان مدل آمیخته‌ی لوجیتی چندجمله‌ای با اثرهای تصادفی ناحیه‌ی ویژه استفاده می‌شود که این مدل قبلاً معرفی شده است که با استفاده از آن براوردهای مدل‌مبنای هم‌زمان برای نسبت‌های بیکاری و اشتغال فراهم می‌گردد.

مفهوم‌های برآورد کوچک‌ناحیه‌ای و دو مدل رگرسیون لوژیستیک و مدل آمیخته‌ی لوجیتی چندجمله‌ای در بخش ۲ معرفی می‌شوند. پس از معرفی طرح نیروی کار مرکز آمار ایران در بخش ۳، در یک مطالعه‌ی شبیه‌سازی و با استفاده از داده‌های نیروی کار فصل اول سال ۱۳۸۹ در بخش ۴، نرخ بیکاری در استان‌های کشور تحت دو مدل رگرسیون لوژیستیک و مدل آمیخته‌ی لوجیتی چندمتغیره را برآورد کرده و آن‌ها را با معیار میانگین توان دوم خطا با هم مقایسه می‌کنیم.

۲- برآورد کوچک‌ناحیه‌ای

هرگاه اندازه‌ی نمونه‌ای در زیرگروه‌های مورد نظر از جامعه بسیار کوچک یا صفر باشد، برآوردهای مستقیم دقت کافی را ندارند. در چنین حالتی، استفاده از رهیافت مدل‌مبنا

منجر به ساختن برآوردهای نامستقیم و مدل‌مبنایی می‌شود که در مقایسه با برآوردهای مستقیم طرح‌مبنا دقیق‌تر هستند.

مدل‌های آماری، مسیری را که داده‌های مرتبط در فرایند برآورد شرکت می‌کنند، نشان می‌دهند. استفاده‌ی مؤثر از برآوردهای پارامترهای جامعه‌ای به کیفیت مدل برازش داده‌شده بستگی دارد. مدل‌های کوچک‌ناحیه‌ای با استفاده از داده‌های تکمیلی مثل داده‌های سرشماری یا داده‌های ثبتي، ناحیه‌ها یا دوره‌های زمانی مختلف را به یکدیگر پیوند می‌دهد. در صورت وجود متغیرهای کمکی، که از داده‌های سرشماری یا آمارگیری‌های بزرگ قبلی به دست می‌آیند، می‌توان مدل‌های کوچک‌ناحیه‌ای را بر اساس نوع اطلاعات در دسترس به دو گروه گسترده رده‌بندی کرد: اول، مدل‌های در سطح ناحیه که میانگین‌های کوچک‌ناحیه‌ای را به متغیرهای کمکی ناحیه‌ویژه ربط می‌دهند. اگر داده‌ها در سطح واحد آماری در دسترس نباشند، استفاده از این مدل‌ها ضروری می‌شوند. دوم، مدل‌های در سطح واحد آماری که مقادیرهای متغیر پاسخ واحد آماری را به متغیرهای کمکی واحدویژه ربط می‌دهند [۱].

۱-۲- مدل‌های کوچک‌ناحیه‌ای برای برآورد نرخ بیکاری

۱-۱-۲- مدل رگرسیون لوژستیکی

فرض کنید متغیرهای z_i مربوط به واحد z ام در کوچک‌ناحیه‌ی z ام با توجه به این که ویژگی خاصی را دارند یا نه، مقدارهای ۱ و ۰ را بپذیرند، اگر علاقه‌مند به برآورد نسبت در کوچک‌ناحیه باشیم؛ آن‌گاه با یاری گرفتن از متغیرهای کمکی و تعریف مدل مناسبی که متغیرهای کمکی را به متغیر پاسخ ربط دهد؛ می‌توانیم برآوردگری مدل‌مبنا به دست آوریم که ضعف خصوصیات برآوردگر مستقیم نسبت در کوچک‌ناحیه را جبران کند. مدل رگرسیون لوژستیکی با اثرهای تصادفی ناحیه‌ای، مدل مناسبی برای برآورد پارامتر نسبت است. مک‌گیون و تامبرلن در سال ۱۹۸۹ برای برآورد نسبت این مدل را پیش‌نهاد دادند [۳]. در این مدل متغیرهای کمکی از طریق یک مدل لججیتی با متغیر پاسخ ارتباط برقرار می‌کنند و معمولاً فرض می‌شود که اثرهای تصادفی حاصل از تغییرات ناحیه‌ها در مدل، از توزیع نرمال پیروی می‌کنند.

در این مقاله با استفاده از داده‌های آمارگیری از نیروی کار فصل اول سال ۱۳۸۹ مرکز آمار ایران می‌توان نسبت‌های بیکاری و اشتغال، مجموع بیکاران و شاغلان در هر استان را با استفاده از دو مدل مجزای رگرسیون لوژیستیکی (یک مدل برای برآورد نسبت بیکاری و یک مدل برای برآورد نسبت اشتغال در استان‌ها) برآورد کرد. مشکل استفاده از این مدل‌ها این است که ممکن است نسبت‌های برآورد شده درون دامنه‌ی $[0, 1]$ نباشند یا مجموع دو نسبت بیش از یک شود یعنی برآورد شاغلان به علاوه‌ی برآورد بیکاران از مجموع کل جامعه بیشتر شود. مشکل دیگر این است که این مدل‌ها همبستگی قوی موجود بین مجموع شاغلان و بیکاران در هر استان را در مسیر برآورد لحاظ نمی‌کنند. فرض کنید واحدهای آماری هر استان i در رده‌های سنی-جنسیتی z گروه‌بندی می‌شوند و به شرط معلوم بودن p_{ij} ها، مقادیرهای متغیر پاسخ y_{ijl} (به ازای $l = 1, \dots, N_{ij}$ که N_{ij} اندازه‌ی جامعه‌ی استان i و رده‌ی سنی-جنسیتی z است) در خانه‌ی (i, z) جدول متقاطع، متغیرهای برنولی مستقل با احتمال مشترک p_{ij} باشند.

y_{ijl_1} و y_{ijl_2} به ترتیب متغیر کمکی مربوط به وضعیت بیکاری و اشتغال فرد و p_{ij_1} و p_{ij_2} به ترتیب احتمال‌های بیکاری و اشتغال در رده‌ی سنی-جنسیتی z و استان i ($i = 1, \dots, 30$) باشند. دو مدل رگرسیون لوژیستیکی (برای شاغلان و برای بیکاران) به صورت زیر برآزش داده شدند:

$$y_{ijkl} | p_{ijk} \sim \text{Bernoulli}(p_{ijk})$$

$$\text{Logit}(p_{ijk}) = \log \frac{p_{ijk}}{1-p_{ijk}} = X_{ij} \beta_k + v_i,$$

$$(1) \quad i = 1, \dots, 30, \quad j = 1, \dots, 6, \quad k = 1, 2$$

که در آن X_{ij} ماتریس X_{ij} معلوم از متغیرهای کمکی (سطح سواد و رده‌های سنی-جنسیتی که در بخش ۳ معرفی خواهند شد)، β_k بردار ضرایب‌های رگرسیونی و $v_i \sim N(0, \sigma_v^2)$ اثرهای تصادفی ناحیه‌ای مربوط به استان i است.

چون فراوانی افراد در رده‌ی سوم متغیر سطح سواد (کارشناسی ارشد و دکتری) بسیار کم بود، گروه دوم و سوم متغیر سطح سواد (کارشناسی ارشد و دکتری با کارشناسی و کاردانی) با یکدیگر ادغام شدند. بنا بر این متغیر کمکی سطح سواد در ۳ رده و متغیر

کمکی سنی-جنسیتی در ۶ رده وارد مدل شدند و آخرین رده از هر متغیر به عنوان رده‌ی پایه در نظر گرفته شد.

می‌توان با استفاده از برآورد ضریب‌های رگرسیونی، β_k و مقادیرهای پیش‌گویی شده‌ی اثرهای تصادفی، \hat{v}_i ، احتمال‌های بیکاری و اشتغال را با استفاده از عبارت زیر برای رده‌های سنی-جنسیتی و استان برآورد کرد (زبرنویس L به مدل رگرسیون لوژستیکی اشاره دارد):

$$\hat{p}_{ijk}^L = \frac{\exp(x_{ij}\hat{\beta}_k + \hat{v}_i)}{1 + \exp(x_{ij}\hat{\beta}_k + \hat{v}_i)},$$

$$i = 1, \dots, 30, \quad j = 1, \dots, 6, \quad k = 1, 2.$$

در حالت کلی اگر N_i اندازه‌ی جامعه‌ی کوچک‌ناحیه‌ی i ام y_{ij} و متغیر پاسخ مربوط به واحد ($= 1, \dots, N_i$) j ام در ناحیه‌ی i ($= 1, \dots, m$) باشد که تنها مقادیرهای صفر و یک را می‌پذیرد، رویال نسبت ([۲]) در ناحیه‌ی i ام، $P_i = \sum_{j=1}^{N_i} \frac{y_{ij}}{N_i}$ ، را می‌توان به صورت زیر بازنویسی کرد:

$$P_i = \frac{\sum_{j \in S} y_{ij} + \sum_{j \in S^c} y_{ij}^r}{N_i}$$

که در آن جمع روی $j \in S$ مجموع متغیرهای پاسخ مربوط به واحدهای نمونه‌ای y_{ij} ، و جمع روی $j \in S^c$ مجموع متغیرهای پاسخ واحدهای نانمونه‌ای y_{ij}^r ، در ناحیه‌ی i ام هستند. متغیرهای پاسخ واحدهای نمونه‌ای که از نمونه‌ی مورد نظر معلوم هستند اما متغیرهای پاسخ واحدهای نانمونه‌ای در کوچک‌ناحیه‌ی i ام نامعلوم بوده و با به کارگیری مدل رگرسیون لوژستیکی می‌توان آن‌ها را برآورد کرد.

بنا بر این مجموع بیکاران و شاغلان نانمونه‌ای در رده‌های سنی-جنسیتی و استان با استفاده از مقادیرهای احتمال برآورد شده‌ی بیکاری و اشتغال به صورت زیر برآورد می‌شوند، زیرا بر اساس نظریه‌ی رویال [۲] برآورد مجموع شاغلان و بیکاران در هر استان مستلزم پیش‌گویی مجموع شاغلان و بیکاران نانمونه‌ای در هر استان است.

$$\hat{y}_{ijk}^r = n_{ij}^r \hat{p}_{ijk}^L \quad k = 1, 2$$

که در آن \hat{p}_{ij1}^L و \hat{p}_{ij2}^L به ترتیب برآورد احتمال‌های بیکاری و اشتغال با استفاده از مدل رگرسیون لوژیستیکی و \hat{y}_{ij1}^r و \hat{y}_{ij2}^r برآورد مجموع بیکاران و شاغلان نانمونه‌ای در رده‌ی سنی-جنسیتی j ام و استان i ام هستند. اگر N_{ij} و n_{ij} به ترتیب اندازه‌های جامعه‌ای و نمونه‌ای در رده‌ی سنی-جنسیتی j و استان i ام معلوم باشند، آنگاه $n_{ij}^r = N_{ij} - n_{ij}$ اندازه‌ی نانمونه‌ای در این رده‌ها است. مجموع شاغلان و بیکاران در هر استان را می‌توان با استفاده از عبارت

$$\hat{\delta}_{ik}^L = \sum_{j=1}^6 (y_{ijk} + y_{ij2k}^r), \quad i = 1, \dots, 30, \quad k = 1, 2$$

برآورد کرد، که در آن y_{ij1} و y_{ij2} به ترتیب مجموع بیکاران و شاغلان نمونه‌ای و \hat{y}_{ij1}^r و \hat{y}_{ij2}^r مجموع بیکاران و شاغلان نانمونه‌ای برآوردشده در رده‌ی سنی-جنسیتی j ام و استان i ام هستند.

با استفاده از برآورد مجموع بیکاران و شاغلان به دست آمده تحت مدل رگرسیون لوژیستیکی، برآوردهای نرخ بیکاری در استان‌ها با استفاده از رابطه‌ی

$$\hat{ur}_i^L = 100 \times \frac{\hat{\delta}_{i1}^L}{\hat{\delta}_{i1}^L + \hat{\delta}_{i2}^L}, \quad i = 1, \dots, 30.$$

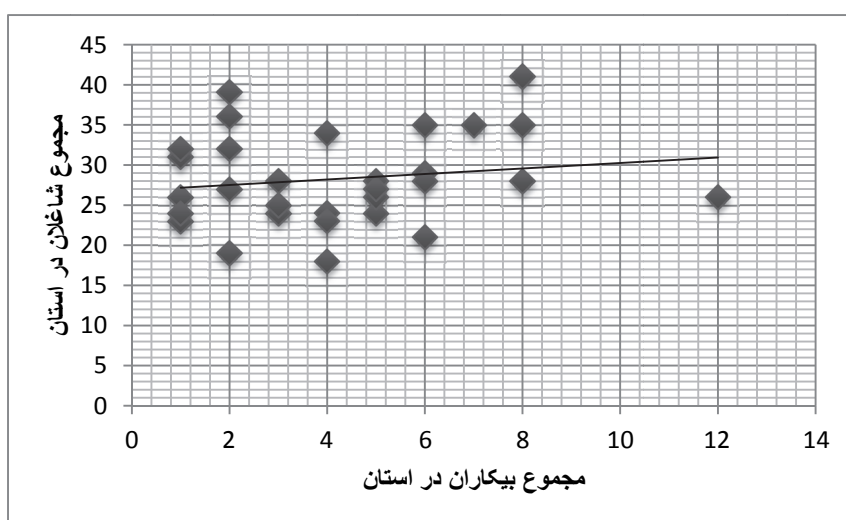
به دست می‌آیند، که در آن $\hat{\delta}_{i1}^L$ و $\hat{\delta}_{i2}^L$ به ترتیب برآورد مجموع بیکاران و شاغلان تحت مدل رگرسیون لوژیستیکی در استان i هستند.

۲-۱-۲- مدل آمیخته‌ی لوجیتی چندجمله‌ای

استفاده از مدل آمیخته‌ی خطی یک ابزار عمومی در برآوردهای کوچک‌ناحیه‌ای است. اثرهای تصادفی ناحیه‌ای در این مدل‌ها تغییرات بین ناحیه‌ها و همبستگی بین واحدهای درون یک ناحیه را (که توسط متغیرهای کمکی بیان نمی‌شوند) نشان می‌دهند. در کاربرد حاضر در آمارگیری نیروی کار، مجموع بیکاران و شاغلان در هر استان را می‌توان از طریق دو مدل جداگانه‌ی رگرسیون لوژیستیکی (یک مدل برای برآورد نسبت بیکاری و

یک مدل برای برآورد نسبت اشتغال در استان‌ها) برآورد کرد اما یک اشکال این مدل‌ها این است که همبستگی قوی موجود بین نسبت‌های بیکاران، شاغلان و غیر فعالان اقتصادی را در مدل منظور نمی‌کند.

برای بررسی رابطه‌ی بین مجموع شاغلان و بیکاران در هر ترکیب سنی-جنسیتی و استان در شکل ۱ این دو متغیر را در مقابل هم رسم کردیم.



شکل ۱- مجموع بیکاران در مقابل مجموع شاغلان در استان‌ها

با توجه به نمودار نقطه‌ها در طول یک پهنا با شیب مثبت پراکنده شده‌اند؛ این نمودار نشان می‌دهند که تعداد افراد شاغل و بیکار نمونه‌ای به صورت خطی با یکدیگر ارتباط دارند. تمرکز نقطه‌ها در گوشه‌ی پایینی سمت چپ وجود یک توزیع توأم بسیار چوله را برای این دو متغیر بازگو می‌کند.

روش جدیدی برای برآورد مجموع و نسبت‌های بیکاران و شاغلان و همچنین برآورد نرخ‌های بیکاری در کوچک‌ناحیه‌ها توسط مولینا و همکارانش معرفی شده است [۴]. در این روش مجموع بیکاران و شاغلان در کوچک‌ناحیه‌ها از یک مدل آمیخته‌ی لوجیتی دو متغیره با اثرهای تصادفی ناحیه‌ای پیروی می‌کنند. این مدل همبستگی قوی موجود بین نسبت‌های بیکاران و شاغلان را در نظر می‌گیرد، همچنین برآوردهای مدل مبنای هم‌زمان

را برای مجموع بیکاران و شاغلان در کوچک‌ناحیه‌ها فراهم می‌کند. برآورد مجموع غیر فعالان اقتصادی با کم کردن از کل جمعیت محاسبه می‌شود. در کاربرد حاضر در آمارگیری نیروی کار، فرض کنید y_{ij1} ، y_{ij2} و y_{ij3} به ترتیب مجموع بیکاران، شاغلان و غیر فعالان اقتصادی نمونه‌ای در رده‌ی سنی-جنسیتی j و استان i ، $n_{ij} = y_{ij1} + y_{ij2} + y_{ij3}$ اندازه‌ی نمونه‌ای رده‌ی سنی-جنسیتی j در استان i و p_{ij1} ، p_{ij2} و p_{ij3} به ترتیب احتمال‌های بیکاری و اشتغال و غیرفعال اقتصادی بودن در رده‌ی سنی-جنسیتی j و استان i باشند. فرض می‌کنیم بردارهای $(y_{ij1}$ و y_{ij2} و $y_{ij3})$ به شرط معلوم بودن v_i (اثر تصادفی ناحیه‌ای) و n_{ij} مستقل و دارای توزیع چندجمله‌ای با تابع چگالی زیر هستند:

$$(2) \quad f(y_{ij1}, y_{ij2} | v_i) = \frac{n_{ij}!}{y_{ij1}! y_{ij2}! y_{ij3}!} p_{ij1}^{y_{ij1}} p_{ij2}^{y_{ij2}} p_{ij3}^{y_{ij3}}.$$

همچنین فرض می‌کنیم احتمال‌های p_{ij1} و p_{ij2} با متغیرهای کمکی سطح سواد و رده‌های سنی-جنسیتی و اثرهای تصادفی ناحیه‌ای دارای رابطه‌ی زیر هستند:

$$(3) \quad \log\left(\frac{p_{ijk}}{p_{ijr}}\right) = X_{ij}\beta_k + v_i, \\ i = 1, \dots, 30, \quad j = 1, \dots, 6, \quad k = 1, \dots, 2$$

که در آن β_k ($= 1, 2$) بردار ضریب‌های رگرسیونی، X_{ij} ماتریس معلوم از متغیرهای کمکی وضع سواد و رده‌های سنی-جنسیتی هستند. در این مدل رده‌ی غیر فعالان اقتصادی را به‌عنوان رده‌ی پایه در نظر گرفتیم، همچنین فرض کردیم که $v_i \sim \text{indN}(0, \sigma_v^2)$. مدل‌های (۱) و (۲) به مجموع شاغلان و بیکاران نمونه‌ای استان‌ها (به‌صورت توأم) برازش داده می‌شوند. پس از برآورد ضریب‌های رگرسیونی و پیش‌گویی اثرهای تصادفی ناحیه‌ای، احتمال‌های بیکاری و اشتغال در رده‌های سنی-جنسیتی و استان از طریق عبارت

$$\hat{p}_{ijk}^M = \frac{\exp(X_{ij}\hat{\beta}_k + \hat{v}_i)}{1 + \sum_{k=1}^2 \exp(X_{ij}\hat{\beta}_k + \hat{v}_i)}$$

$$i = 1, \dots, 30, \quad j = 1, \dots, 6, \quad k = 1, 2$$

برآورد می‌شوند، که در آن β_k بردار ضریب‌های رگرسیونی، X_{ij} ماتریس معلوم از متغیرهای کمکی وضع سواد و رده‌های سنی-جنسیتی و \hat{V}_i مقدار پیش‌گویی‌شده‌ی اثر تصادفی ناحیه‌ی (استان) i ام است (زبرنویس M به مدل آمیخته‌ی لوجیتی چندجمله‌ای اشاره دارد).

با استفاده از احتمال‌های برآوردشده‌ی بیکاری و اشتغال، مجموع بیکاران و شاغلان نانمونه‌ای در رده‌ی سنی-جنسیتی j ام و استان i ام به‌صورت زیر پیش‌گویی می‌شوند:

$$\hat{y}_{ijk}^r = n_{ij}^r \hat{\rho}_{ijk}^M$$

که در آن n_{ij}^r اندازه‌ی نانمونه‌ای برای رده‌ی سنی-جنسیتی j ام و استان i ام است. در نهایت مجموع شاغلان و بیکاران در هر استان با استفاده از عبارت زیر برآورد می‌شوند:

$$\hat{\delta}_{ik}^M = \sum_{j=1}^6 (y_{ijk} + \hat{y}_{ijk}^r), \quad i = 1, \dots, 30, \quad j = 1, \dots, 6, \quad k = 1, 2$$

که در آن y_{ij1} و y_{ij2} به‌ترتیب مجموع بیکاران و شاغلان نمونه‌ای و \hat{y}_{ij1}^r و \hat{y}_{ij2}^r به‌ترتیب مقادیرهای پیش‌گویی‌شده‌ی بیکاران و شاغلان نانمونه‌ای در رده‌های سنی-جنسیتی j ام و استان i هستند.

با استفاده از مجموع شاغلان و بیکاران برآوردشده تحت مدل لوجیتی آمیخته‌ی چندجمله‌ای برآورد نرخ بیکاری برای استان‌ها با استفاده از عبارت زیر به دست می‌آید:

$$\hat{ur}_i^M = 100 \times \frac{\hat{\delta}_{i1}^M}{\hat{\delta}_{i1}^M + \hat{\delta}_{i2}^M}, \quad i = 1, \dots, 30, \quad j = 1, \dots, 6, \quad k = 1, 2$$

که در آن $\hat{\delta}_{i1}^M$ و $\hat{\delta}_{i2}^M$ برآورد مجموع بیکاران و شاغلان در استان‌ها تحت مدل آمیخته‌ی لوجیتی چندجمله‌ای هستند.

۳- طرح آمارگیری نیروی کار

طرح آمارگیری نیروی کار بر اساس نیاز برنامه‌ریزان با هدف محاسبه‌ی شاخص‌های فصلی و سالانه‌ی نیروی کار از جمله نرخ مشارکت اقتصادی، نرخ بیکاری و تغییرات آن‌ها طراحی شده است. این طرح منطبق بر آخرین توصیه‌های سازمان بین‌المللی کار است و امکانی را برای مقایسه‌ی بین‌المللی شاخص‌های نیروی کار فراهم می‌کند. در این طرح برای بهبود برآورد تغییرات، از روش نمونه‌گیری چرخشی استفاده می‌شود. برای دستیابی به برآوردهای فصلی با توجه به ماهیت اقلام و پرسش‌های پرسش‌نامه، آمارگیری در ماه میانی هر فصل در دو هفته‌ی مشخص انجام می‌شود و برخی پرسش‌های پرسش‌نامه در مورد هفته‌ی قبل از آن دو هفته (هفته‌های مرجع) پرسیده می‌شود. همچنین به دلیل ماهیت ادواری طرح، برای طراحی بهینه‌ی آن، از نمونه‌ی پایه استفاده می‌شود. نمونه‌ی پایه نمونه‌ای است که می‌توان از آن برای چند دوره از یک آمارگیری ادواری یا چند آمارگیری، زیرنمونه‌هایی را انتخاب کرد. نمونه‌ی پایه‌ی این طرح، طی فرایندی از چارچوب سرشماری عمومی نفوس و مسکن سال ۱۳۸۵ و تقسیمات کشوری سال ۱۳۸۶ تهیه و مورد استفاده قرار گرفته است.

آمارگیری در این طرح به صورت نمونه‌گیری است و اطلاعات مربوط به خانوار از طریق مصاحبه‌ی حضوری مأمور آمارگیر با مطلع‌ترین عضو خانوار و تکمیل پرسش‌نامه گردآوری می‌شود. اطلاعات هر یک از اعضای خانوار نیز از طریق مصاحبه‌ی حضوری مأمور آمارگیر با هر یک از اعضای ده ساله و بیش‌تر خانوار گردآوری می‌شود.

برای دستیابی به برآوردهای مورد نظر، ابتدا باید داده‌های مربوط به هر یک از خانوارهای نمونه و اعضای آن‌ها وزن‌دهی و سپس از فرمول‌های برآورد استفاده شود. وزن‌دهی در سه مرحله‌ی «اعمال وزن پایه»، «تعدیل وزن برای بی‌پاسخی واحد» و «تعدیل وزن بر اساس پیش‌بینی‌های جمعیتی» انجام می‌گیرد. وزنی که پس از اعمال این مرحله‌ها به دست می‌آید، تعیین می‌کند که هر فرد در نمونه نماینده‌ی چند نفر در جامعه است. این آمارگیری با هدف ارزیابی برآوردها در سطح استان‌های کشور طراحی شده است.

این مقاله داده‌های طرح آمارگیری از نیروی کار مرکز آمار ایران به تعداد ۱۲۷۱۹۸ فرد در فصل اول سال ۱۳۸۹ به عنوان جامعه‌ی هدف در نظر گرفته شده است. با استفاده از برآورد واریانس نسبت بیکاری در فصل اول سال ۱۳۸۹ اندازه‌ی نمونه‌ی بهین در سطح

کشور را در طرح نمونه‌گیری تصادفی ساده به دست آوردیم. چون کشور به استان‌ها افراز شده است بنا بر این اگر واحدها با استفاده از طرح نمونه‌گیری طبقه‌بندی انتخاب شوند دقت برآوردها افزایش می‌یابد. بنا بر این با در نظر گرفتن اثر طرح برابر با $\frac{1}{8}$ اندازه‌ی نمونه‌ای بهین در سطح کشور در طرح نمونه‌گیری طبقه‌بندی به دست آمد. با یکسان فرض کردن واریانس نسبت بیکاری و هزینه‌ی آمارگیری در استان‌ها اندازه‌های نمونه‌ای قرار گرفته در استان‌ها را با استفاده از تخصیص متناسب به دست آوردیم.

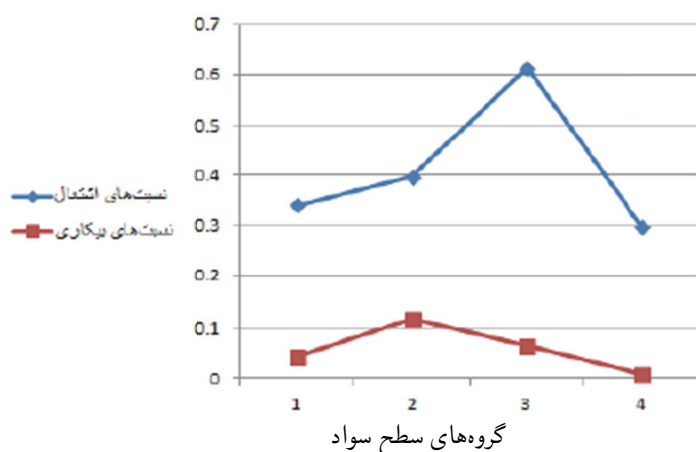
در انتخاب متغیرهای کمکی و به کارگیری آن‌ها در مدل‌های مورد نظر با محدودیت‌های بسیاری مواجه بودیم، زیرا هر چه سطح کوچک‌ناحیه‌ها از استان‌ها کوچک‌تر می‌شوند متغیرهای کمکی محدود می‌شوند. بنا بر این با توجه به پرسش‌نامه متغیرهای کمکی سطح سواد و رده‌های سنی-جنسیتی به‌عنوان متغیرهای کمکی در برازش مدل‌های رگرسیون لوژیستیک و آمیخته‌ی لجیتی چندجمله‌ای برای برآورد مجموع بیکاران و شاغلان استفاده شدند. لذا متغیر سطح سواد در ۴ رده و متغیر کمکی سنی-جنسیتی در ۶ رده وارد مدل شدند. نمادهای رده‌های متغیرهای کمکی در جدول ۱ تعریف شده‌اند.

جدول ۱- توضیح رده‌های متغیرهای کمکی

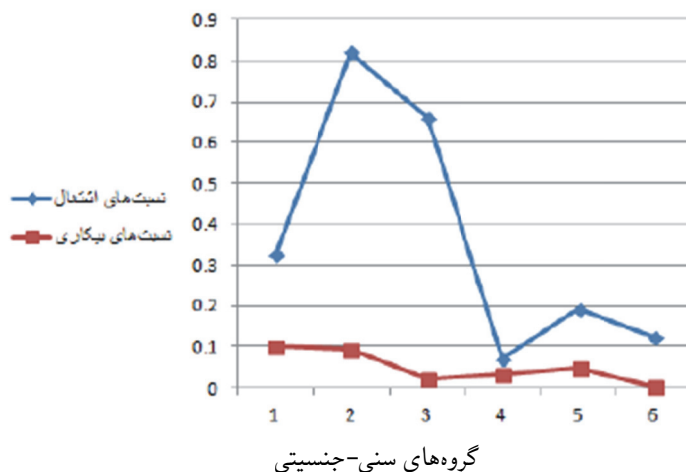
متغیر	رده	دامنه
سن-جنسیتی	۱	مردان [۱۰,۲۵] سال
	۲	مردان (۲۵,۴۰) سال
	۳	مردان بالای ۴۰ سال
	۴	زنان [۱۰,۲۵] سال
	۵	زنان (۲۵,۴۰) سال
	۶	زنان بالای ۴۰ سال
سطح سواد	۱	دیپلم و زیر دیپلم
	۲	کاردانی و کارشناسی
	۳	کارشناسی ارشد و دکتری
	۴	بی‌سواد

در شکل‌های ۲ و ۳ با رسم میانگین نسبت‌های بیکاری و اشتغال (روی کل کوچک‌ناحیه‌ها)، در هر یک از گروه‌های سنی-جنسیتی و سطح سواد توان این متغیرها را در پیش‌گویی متغیر پاسخ بررسی می‌کنیم. مشاهده می‌کنیم که میانگین نسبت‌ها در بین

گروه‌ها برای دو متغیر تغییر می‌کنند. این تغییرات در گروه‌های بیکاری و اشتغال موازی نیستند. به نظر می‌رسد که شاخص‌های گروه‌ها در هر یک از متغیرها در پیش‌گویی نسبت‌های بیکاری و اشتغال به صورت بالقوه‌ای مفید هستند.



شکل ۲- نمودار میانگین نسبت‌های بیکاری و اشتغال (روی کوچک‌ناحیه‌ها) در گروه‌های سطح سواد



شکل ۳- نمودار میانگین نسبت‌های بیکاری و اشتغال (روی کوچک‌ناحیه‌ها) در گروه‌های سنی-جنسیتی

با استفاده از تحلیل واریانس فرض صفر را مبنی بر این که میانگین نسبت‌های بیکاری و اشتغال در ۶ گروه سنی-جنسیتی و ۴ گروه سطح سواد با یکدیگر اختلاف معناداری دارند یا خیر آزمون کردیم. نتیجه‌ی آزمون نشان می‌دهد که گروه‌بندی‌های سنی-جنسیتی و سطح سواد در برآورد نسبت‌های بیکاری و اشتغال تأثیرگذار هستند و دقت برآوردها با وام گرفتن از این متغیرهای کمکی افزایش خواهد یافت.

۴- مطالعه‌ی شبیه‌سازی

به منظور مقایسه‌ی برآوردهای نرخ‌های بیکاری حاصل از مدل‌های رگرسیون لوژیستیکی و مدل آمیخته‌ی لوجیتی چندجمله‌ای با هم و با برآوردهای مستقیم یک مطالعه‌ی شبیه‌سازی انجام شد.

۱۰۰۰ نمونه‌ی خودگردان ۲۴۹۳ تایی برای ۳۰ کوچک‌ناحیه (استان‌ها) تولید شده است. برای هر نمونه‌ی خودگردان دو مدل رگرسیون لوژیستیکی و مدل آمیخته‌ی لوجیتی چندجمله‌ای به داده‌ها برازش داده شدند. برآوردهای نرخ‌های بیکاری تحت مدل‌های رگرسیون لوژیستیکی و مدل آمیخته‌ی لوجیتی چندجمله‌ای و MSE آن‌ها محاسبه و در نهایت عمل‌کرد برآوردهای مستقیم و برآوردهای مدل‌مبنا با یکدیگر مقایسه شدند.

۴-۱- تعیین اندازه‌ی نمونه‌ی اصلی

برای تعیین اندازه‌ی نمونه‌ی اصلی در سطح کشور با استفاده از نسبت بیکاری، با توجه به آمارهای منتشرشده توسط مرکز آمار ایران نرخ بیکاری در فصل اول سال ۱۳۸۹ که برای جمعیت ده ساله و بیش‌تر برابر با ۱۳/۵ درصد بوده، واریانس نرخ بیکاری را به صورت تقریبی برابر با ۰/۱۶۸ در نظر گرفتیم.

اندازه‌ی نمونه در سطح کشور در طرح نمونه‌گیری تصادفی ساده به صورت

$$n_{srs} = \frac{z_{\alpha/2}^2 s^2}{e^2 + \frac{z_{\alpha/2}^2 s^2}{N}}$$

به دست آورده شده است که در آن $\alpha = ۰/۰۵$ ، $Z_{\alpha} = ۱/۹۶$ ، $N=۱۲۷۱۹۸$ و حاشیه‌ی خطای تقریبی $e = ۰/۰۱۲$ است.

چون جامعه به طبقه‌هایی (استان‌ها) افراز شده است بنا بر این اگر واحدها با استفاده از طرح نمونه‌گیری طبقه‌ای از جامعه انتخاب شوند دقت برآوردها در مقایسه با طرح نمونه‌گیری تصادفی ساده افزایش می‌یابد. لذا برای این که با همین دقت و با استفاده از طرح نمونه‌گیری طبقه‌ای پارامترها را برآورد کنیم اثر طرح را برابر با $deff = ۰/۸$ در نظر گرفتیم و اندازه‌ی نمونه‌ی طبقه‌ای در سطح کشور را با استفاده از عبارت زیر به دست آوردیم:

$$n_{sts} = n_{srs} \times deff$$

سپس با استفاده از نمونه‌ی کشوری اندازه‌ی نمونه‌های قرار گرفته در استان‌ها را با استفاده از تخصیص متناسب از طریق رابطه‌ی زیر به دست آوردیم (با توجه به داده‌های در دست، واریانس نسبت بیکاری و همچنین هزینه‌ی آمارگیری در استان‌ها را می‌توان تقریباً برابر گرفت).

$$n_i = \frac{N_i}{N} n_{sts} \quad i = ۱, \dots, ۳۰$$

که در آن N_i اندازه‌ی جامعه‌ای استان i و N اندازه‌ی جامعه‌ای کشور و n_{sts} اندازه‌ی نمونه‌ای با استفاده از طرح نمونه‌گیری طبقه‌ای در سطح کشور هستند. اندازه‌های جامعه‌ای، N_i ، اندازه‌ی نمونه‌ای قرار گرفته در استان‌ها، n_i ، به همراه اندازه‌ی نمونه‌ای بهین n_i^{opt} در جدول ۲ آورده شده است.

با مقایسه‌ی اندازه‌های نمونه‌ای بهین و اندازه‌های نمونه‌های قرار گرفته در استان‌ها به این نتیجه می‌رسیم که می‌توان استان‌ها را به‌عنوان کوچک‌ناحیه در نظر گرفت و از فن‌های برآورد کوچک‌ناحیه‌ای برای برآورد پارامترهای مورد نظر بهره گرفت.

جدول ۲- اندازه‌ی جامعه‌ای و نمونه‌ای بهین به همراه اندازه‌های نمونه‌های قرار گرفته در هر استان از نمونه‌ی کل

$n_{i,opt}$	n_i	N_i	استان	کد استان
۵۲۳	۸۰	۴۰۸۸	مرکزی	۰۰
۲۶۱	۸۶	۳۵۰۲	گیلان	۰۱
۴۸۸	۸۰	۴۱۵۶	مازندران	۰۲
۴۲۸	۹۳	۴۷۴۷	آذربایجان شرقی	۰۳
۴۳۰	۹۱	۴۶۶۸	آذربایجان غربی	۰۴
۶۸۴	۱۰۵	۵۳۷۵	کرمانشاه	۰۵
۵۶۷	۸۸	۴۴۹۶	خوزستان	۰۶
۶۰۶	۱۰۴	۵۳۴۳	فارس	۰۷
۲۳۷	۷۲	۳۷۲۲	کرمان	۰۸
۴۳۰	۹۱	۴۶۵۰	خراسان رضوی	۰۹
۲۹۳	۷۶	۳۸۹۵	اصفهان	۱۰
۵۵۲	۸۸	۴۴۹۱	سیستان و بلوچستان	۱۱
۲۹۹	۹۴	۴۸۲۱	کردستان	۱۲
۴۸۲	۸۲	۴۲۸۲	همدان	۱۳
۵۱۸	۸۹	۴۵۹۷	چهارمحال بختیاری	۱۴
۶۵۶	۱۰۰	۵۱۲۷	لرستان	۱۵
۵۲۱	۷۷	۳۹۷۵	ایلام	۱۶
۴۹۰	۸۳	۴۳۰۹	کهگیلویه و بویر احمد	۱۷
۴۹۹	۸۲	۴۱۸۷	بوشهر	۱۸
۲۲۶	۷۵	۳۸۵۰	زنجان	۱۹
۲۸۷	۶۳	۳۲۳۰	سمنان	۲۰
۲۷۴	۷۶	۳۸۸۵	یزد	۲۱
۵۹۵	۹۳	۴۷۶۷	هرمزگان	۲۲
۴۰۳	۸	۴۵۱۰	تهران	۲۳
۲۳۶	۷۳	۳۷۶۷	اردبیل	۲۴
۲۵۷	۶۶	۳۴۰۵	قم	۲۵
۲۵۱	۷۳	۳۷۵۴	قزوین	۲۶
۵۸۳	۸۲	۴۲۱۵	گلستان	۲۷
۴۳۳	۸۱	۴۱۶۱	خراسان شمالی	۲۸
۱۰۰	۶۲	۳۲۱۲	خراسان جنوبی	۲۹
۱۲۶۰۹	۲۴۹۳	۱۲۷۱۹۸	کل	

۱-۴- برآوردهای مستقیم پارامترهای نیروی کار

فرض کنید n_i و N_i به ترتیب اندازه‌ی نمونه‌ای و جامعه‌ای در استان i ام و y_{ij1} و y_{ij2} به ترتیب تعداد بیکاران و شاغلان در نمونه‌های خودگردان رده‌ی سنی-جنسیتی j ($= 1, \dots, 6$) و استان i ($= 1, \dots, 30$) باشند، برآوردهای مستقیم نسبت‌های بیکاری و اشتغال در استان i ام در نمونه‌های خودگردان ($b = 1, \dots, 1000$) با استفاده از رابطه‌های زیر به دست آمدند:

$$\hat{p}_{ib1}^{dir} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij1},$$

$$\hat{p}_{ib2}^{dir} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij2}, \quad i = 1, \dots, 30$$

برآورد مستقیم خودگردان نسبت‌های بیکاری و اشتغال به ترتیب عبارت‌اند از:

$$\hat{p}_{iB1}^{dir} = \frac{1}{1000} \sum_{b=1}^{1000} \hat{p}_{ib1}^{dir},$$

$$\hat{p}_{iB2}^{dir} = \frac{1}{1000} \sum_{b=1}^{1000} \hat{p}_{ib2}^{dir}, \quad i = 1, \dots, 30$$

با استفاده از برآورد نسبت‌ها، برآوردهای مستقیم مجموع بیکاران و شاغلان به صورت

$$\hat{\delta}_{ib1}^{dir} = N_i \hat{p}_{iB1}^{dir}, \quad \hat{\delta}_{ib2}^{dir} = N_i \hat{p}_{iB2}^{dir}, \quad i = 1, \dots, 30$$

حاصل می‌شوند. بنا بر این می‌توان برآورد مستقیم نرخ‌های بیکاری در استان‌ها را به صورت

$$(۴) \quad \hat{u}_{iB}^{dir} = 100 \times \frac{\hat{\delta}_{ib1}^{dir}}{\hat{\delta}_{ib1}^{dir} + \hat{\delta}_{ib2}^{dir}}, \quad i = 1, \dots, 30$$

به دست آورد. می‌توان میانگین توان دوم خطا برای برآوردهای مستقیم نرخ‌های بیکاری حاصل از نمونه‌های خودگردان را به صورت

$$(۵) \quad M\hat{S}E(\hat{u}r_i^{dir}) = \frac{1}{1000} \sum_{b=1}^{1000} (\hat{u}r_{ib}^{dir} - \hat{u}r_{iB}^{dir})^2, \quad i = 1, \dots, 30$$

برآورد کرد که در آن $\hat{u}r_{ib}^{dir}$ برآوردهای مستقیم نرخ بیکاری در نمونه‌ی خودگردان b ام و $\hat{u}r_{iB}^{dir}$ برآورد خودگردان نرخ بیکاری در استان i ام هستند.

۲-۴- برآزش مدل رگرسیون لوژیستیکی به داده‌ها

برای هر نمونه‌ی خودگردان، دو مدل رگرسیون لوژیستیکی مجزا به صورت مدل (۱) به مجموع بیکاران و شاغلان در رده‌های سنی-جنسیتی و استان‌ها برآزش داده شدند. سپس برآورد مجموع بیکاران و شاغلان تحت مدل رگرسیون لوژیستیکی یعنی $\hat{\delta}_{ib1}^L$ و $\hat{\delta}_{ib2}^L$ به دست آمدند. نرخ‌های بیکاری در استان‌ها، $\hat{u}r_{ib}^L$ ، با استفاده از رابطه‌ی (۴) برآورد شدند (زبرنویس L به مدل رگرسیون لوژیستیکی اشاره دارد). همچنین میانگین توان دوم خطا برای نرخ‌های بیکاری تحت مدل‌های رگرسیون لوژیستیکی، $M\hat{S}E(\hat{u}r_i^L)$ ، با استفاده از رابطه‌ی (۵) تقریب زده شدند.

۳-۴- برآزش مدل آمیخته‌ی لوجیتی چندجمله‌ای به داده‌ها

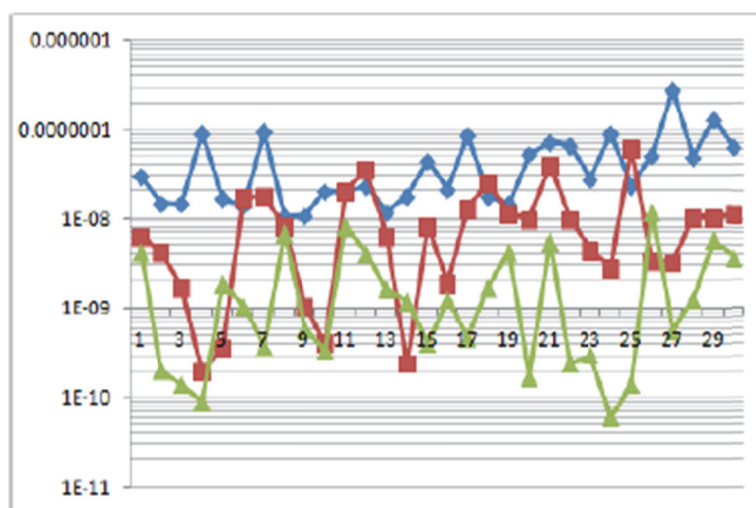
با فرض این که y_{ij1} ، y_{ij2} و y_{ij3} به ترتیب مجموع بیکاران، شاغلان و غیر فعالان اقتصادی در نمونه‌های خودگردان در رده‌ی سنی-جنسیتی j و استان i ، $n_{ij} = y_{ij1} + y_{ij2} + y_{ij3}$ اندازه‌ی نمونه‌ی رده‌ی سنی-جنسیتی j در استان i و p_{ij1} ، p_{ij2} و p_{ij3} به ترتیب احتمال‌های بیکاری و اشتغال و غیرفعال اقتصادی بودن در رده‌ی سنی-جنسیتی j در استان i هستند، مدل آمیخته‌ی لوجیتی چندجمله‌ای (۲) و (۳) به نمونه‌های خودگردان برآزش داده شد و نرخ‌های بیکاری تحت مدل آمیخته‌ی لوجیتی چندجمله‌ای، $\hat{u}r_{ib}^M$ ، با استفاده از عبارت (۴) برآورد شدند (زبرنویس M به مدل لوجیتی آمیخته‌ی چندجمله‌ای اشاره دارد). سپس MSE برآوردهای نرخ بیکاری در کوچک‌ناحیه‌ی i ، $M\hat{S}E(\hat{u}r_i^M)$ ، با استفاده از رابطه‌ی (۵) تقریب زده شدند.

برآورد خودگردان MSE ی برآوردهای نرخ‌های بیکاری حاصل از دو مدل رگرسیون لوژیستیکی و مدل آمیخته‌ی لوژیستیکی چندجمله‌ای در جدول ۳ نشان داده شده‌اند.

جدول ۳- برآورد خودگردان MSE ی برآوردهای نرخ‌های بیکاری مدل رگرسیون لوژیستیکی و مدل آمیخته‌ی لوژیستیکی چندجمله‌ای و برآورد مستقیم

$MSE(\hat{ur}_i^M)$	$MSE(\hat{ur}_i^L)$	$MSE(\hat{ur}_i^{dir})$	استان	کد استان
۴/۱۷	۶/۲۵	۲۵/۵۹	مرکزی	۰۰
۰/۲۰	۴/۱۶	۱۴/۵۳	گیلان	۰۱
۰/۱۴	۱/۷۶	۱۴/۵۱	مازندران	۰۲
۰/۰۹	۰/۱۹۶۱	۹۰/۴۴	آذربایجان شرقی	۰۳
۱/۹۱	۰/۳۷	۱۶/۴۸	آذربایجان غربی	۰۴
۱/۰۷	۱۶/۷۰	۱۳/۸۳	کرمانشاه	۰۵
۰/۳۸	۱۷/۳۶	۹۳/۳۱	خوزستان	۰۶
۶/۹۳	۸/۱۳	۱۰/۶۲	فارس	۰۷
۰/۰۶	۱/۰۵	۱۰/۹۳	کرمان	۰۸
۰/۳۴۹۴	۰/۴۱	۲۰/۰۷	خراسان رضوی	۰۹
۸/۱۵	۱۹/۹۳	۱۹/۷۱	اصفهان	۱۰
۴/۰۶	۳۵/۲۴	۲۲/۹۴	سیستان و بلوچستان	۱۱
۱/۶۱	۶/۳۴	۱۱/۹۷	کردستان	۱۲
۱/۲۲	۰/۲۴	۱۷/۸۰	همدان	۱۳
۰/۴۱	۸/۲۹	۴۵/۵۶	چهارمحال بختیاری	۱۴
۱/۲۳	۱/۹	۲۱/۴۰	لرستان	۱۵
۰/۴۸	۱۲/۸۷	۸۶/۳۰	ایلام	۱۶
۱/۷۳	۲۴/۸۴	۱۷/۳۶	کهگیلویه و بویر احمد	۱۷
۴/۱۹	۱۱/۳۴	۱۴/۴۰	بوشهر	۱۸
۰/۱۷	۹/۷۵	۵۲/۵۶	زنجان	۱۹
۵/۵۲	۳۹/۰۵	۷۳/۲۷	سمنان	۲۰
۰/۲۳	۹/۸۷	۶۷/۸۲	یزد	۲۱
۰/۳۰	۴/۴۶	۲۷/۲۴	هرمزگان	۲۲
۰/۰۶	۲/۸۱	۹۱/۹۶	تهران	۲۳
۰/۱۴	۶۱/۴۴	۲۰/۳۱	اردبیل	۲۴
۱۱/۶۲	۳/۳۹	۵۱/۵۵	قم	۲۵
۰/۵۷	۳/۲۳	۲۷/۷۷	قزوین	۲۶
۱/۲۸	۱۰/۵۷	۴۸/۷۲	گلستان	۲۷
۵/۶۲	۱۰/۲۹	۱۳۱/۱۰	خراسان شمالی	۲۸
۳/۷۲	۱۱/۲۳	۶۲/۷۲	خراسان جنوبی	۲۹

شکل ۴ میانگین توان دوم خطای خودگردان برآوردهای مستقیم و مدل مبنای نرخ‌های بیکاری در مقیاس لگاریتمی (برای وضوح بهتر در مقیاس لگاریتمی رسم شده است) نشان می‌دهد. جدول ۳ و شکل ۴ نشان می‌دهند که \widehat{MSE} ی مربوط به برآوردهای مدل مبنای در تمام استان‌ها کوچک‌تر از \widehat{MSE} ی برآوردهای مستقیم است. همچنین \widehat{MSE} ی برآوردهای حاصل از مدل آمیخته‌ی لوجیتی چندجمله‌ای در بسیاری از استان‌ها کوچک‌تر از \widehat{MSE} ی برآوردهای حاصل از مدل رگرسیون لوجستیکی است زیرا این مدل همبستگی بین مجموع بیکاران و شاغلان در استان‌ها را در فرایند برآورد منظور کرده است.



شکل ۴- میانگین توان دوم خطای خودگردان برآوردهای مستقیم و مدل مبنای نرخ‌های بیکاری در مقیاس لگاریتمی استان‌ها؛ میانگین توان دوم خطای برآوردهای خودگردان نرخ بیکاری استان‌ها: مستقیم \blacklozenge ، تحت مدل رگرسیون لوجستیکی \blacktriangle و تحت مدل آمیخته‌ی لوجیتی چندجمله‌ای \blacksquare

۵- بحث و نتیجه‌گیری

در این مقاله پس از معرفی دو مدل رگرسیون لوجستیکی و مدل آمیخته‌ی لوجیتی چندجمله‌ای در یک مطالعه‌ی شبیه‌سازی این دو مدل به داده‌ها برآزش داده شدند. یافته‌های مطالعه‌ی شبیه‌سازی نشان می‌دهد که اولاً برآوردهای مدل مبنای به دلیل یاری گرفتن از متغیرهای کمکی در مقایسه با برآوردهای مستقیم دقت بیشتری دارند. ثانیاً

براوردهای نرخ‌های بیکاری تحت مدل آمیخته‌ی لوجیتی چندمتغیره در مقایسه با براوردهای نرخ‌های بیکاری بر اساس مدل رگرسیون لوزستیکی دقیق‌تر هستند؛ زیرا این مدل همبستگی بین مجموع بیکاران و شاغلان را در مدل منظور می‌کند.

مرجع‌ها

- [1] Rao, J.N.K. (2003). *Small Area Estimation*, John Wiley & Sons, New York.
- [2] Royall, R.M. (1970). On Finite Population Sampling Theory under Certain Linear Regression Models, *Biometrika*, **57**, 377–387.
- [3] Mac Gibbon, B. and Tomberlin, T.J. (1989). Small Area Estimates of Proportions via Empirical Bayes Techniques, *Survey Methodology*, **15**, 237–252.
- [4] Molina, I., Saei, A. and Lombardía, J. (2007). Small Area Estimates of Labour Force Participation under a Multinomial Logit Mixed Model, *J. R. Statist.* **170**, 975–1000.

فائزه عباس‌زاده

کارشناس ارشد آمار

تهران، خیابان شهید بهشتی، نبش احمد قصیر، دانشکده‌ی اقتصاد دانشگاه علامه طباطبایی، گروه آمار.

رایانشانی: faezeh_abbaszadeh@yahoo.com

حمیدرضا نواب‌پور

دکتری آمار

تهران، خیابان شهید بهشتی، نبش احمد قصیر، دانشکده‌ی اقتصاد دانشگاه علامه طباطبایی، گروه آمار.

رایانشانی: h.navvabpour@src.ac.ir