مجله‌ی بررسی‌های آمار رسمی ایران
سال۲۴، شماره‌ی ۲، پاییز و زمستان ۱۳۹۲، صص ۲۰۵– ۲۲۵

براوردهای کوچک‌ناحیه‌ای نرخ بیکاری

فائزه عباس‌زاده* و حمیدرضا نواب‌پور

دانشگاه علامه طباطبایی

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 براوردهای قابل قبول برای پارامترهای بیکاری و اشتغال در سطح کشور هستیم. اداره‌های آمار ملّی با این مشکل مواجه هستند که در نمونه‌گیری هنگامی که اندازه‌ی نمونه‌ای در سطح کشور بهینه می‌شود، اندازه‌های نمونه‌ای به‌منظور براورد در ناحیه‌های کوچک‌تر از کشور، براوردهای 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 براورد کوچک‌ناحیه‌ای برای ساختن براوردهای قابل قبول بهره گرفت. با استفاده از رویکرد مدل‌مبنا که در آن متغیرهای پاسخ از طریق یک مدل با متغیرهای کمکی ارتباط برقرار می‌کنند می‌توان دقّت براوردهای کوچک‌ناحیه‌ای را با وام گرفتن قدرت از این متغیرها بهبود بخشید. در این مقاله دو مدل رگرسیون لوژستیکی و مدل آمیخته‌ی لوجیتی چندجمله‌ای با اثرهای تصادفی برای براورد 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 براوردهای کوچک‌ناحیه‌ای مدل‌مبنا و طرح‌مبنا با یکدیگر مقایسه می‌شوند.

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

---

* نویسنده‌ی عهده‌دار مکاتبات
دریافت: ۱۲۹۲/۱۲/۱۲، پذیرش: ۱۳۹۳/۹/۳.

۱ـ مقدمه

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

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

وجود نمونه‌ای جامع، شامل تمام ناحیه‌ها به‌طوری که اندازه‌ی نمونه‌ای مربوط به هر ناحیه برای پشتیبانی براوردگرهای 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 براوردهای مستقیم مبتنی بر آن‌ها ناکارا هستند. در این حالت برای افزایش کارایی براوردگرها از روش‌های براورد کوچک‌ناحیه‌ای استفاده می‌شود. برای استفاده از فن براورد کوچک‌ناحیه‌ای نیاز به داشتن اطلاعات کمکی موثق است. بنا بر این متغیرهای کمکی یا متغیرهای پیش‌گو نقش بسیار مهمی را در براوردهای کوچک‌ناحیه‌ای ایفا می‌کنند. موفقیت روش‌های مدل‌مبنا به موجود بودن داده‌های کمکی مناسب بستگی دارد.

روش‌های براورد کوچک‌ناحیه‌ای از طریق یک مدل آماری، رابطه‌ای را بین متغیر پاسخ و اطّلاعات کمکی ایجاد می‌کنند که منجر به ساختن براوردهای کوچک‌ناحیه‌ای مبتنی بر مدل یا براوردگر نامستقیم می‌شود. یاری گرفتن از اطّلاعات کمکی، می‌تواند براوردهایی

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

در آمارگیری نیروی کار اگر وضع فعالیت فرد (شاغل یا بیکار) به‌صورت یک متغیر دو حالتی (۰و۱) Y در نظر گرفته شود به گونه‌ای که متغیر کمکی‌ای مثل X وجود داشته باشد که بر احتمال وقوع حالت ۱ آن تأثیرگذار باشد یعنی $P\left(Y=۱\right)=P(X)$؛ آنگاه دیدگاه‌های مختلفی برای تعیین ارتباط بین Y و X وجود دارد. یکی از این دیدگاه‌ها استفاده از تابع رگرسیون لوژستیکی است که در آن تابع لوجیتی $\log\left(p(x)/۱-p(x)\right)=\log\mathrm{it}(p(x))$ با یک تابع خطی از متغیرهای کمکی و اثرهای ثابت مرتبط است.

با استفاده از برازش دو تابع مجزای رگرسیون لوژستیکی 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 فعّال 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 براورد کوچک‌ناحیه‌ای و دو مدل رگرسیون لوژستیکی 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 لوژستیکی و مدل آمیخته‌ی لوجیتی چندمتغیره را براورد کرده و آن‌ها را با معیار میانگین توان دوم خطا با هم مقایسه می‌کنیم.

۲- براورد کوچک‌ناحیه‌ای

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

منجر به ساختن براوردهای نامستقیم و مدل‌مبنایی می‌شود که در مقایسه با براوردهای مستقیم طرح‌مبنا دقیق‌تر هستند.
مدل‌های آماری، مسیری را که داده‌های مرتبط در فرایند براورد شرکت می‌کنند، نشان می‌دهند. استفاده‌ی مؤثر از براوردهای پارامترهای جامعه‌ای به کیفیت مدل برازش داده‌شده بستگی دارد. مدل‌های کوچک‌ناحیه‌ای با استفاده از داده‌های تکمیلی مثل داده‌های سرشماری یا داده‌های ثبتی، 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 مقدارهای متغیر پاسخ واحد آماری را به متغیرهای کمکی واحدویژه ربط می‌دهند [۱].

۱-۲- مدل‌های کوچک‌ناحیه‌ای برای براورد نرخ بیکاری

۱-۱-۲- مدل رگرسیون لوژستیکی

فرض کنید متغیرهای y_{ij} مربوط به واحد jام در کوچک‌ناحیه‌ی iام با توجه به این که ویژگی خاصی را دارند یا نه، مقدارهای ۱ و ۰ را بپذیرند، اگر علاقه‌مند به براورد 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 براوردگری مدل‌مبنا به دست آوریم که ضعف خصوصّیات براوردگر مستقیم نسبت در کوچک‌ناحیه را جبران کند. مدل رگرسیون لوژستیکی با اثرهای تصادفی ناحیه‌ای، مدل مناسبی برای براورد پارامتر نسبت است. مک‌گیبون و تامبرلن در سال ۱۹۸۹ برای براورد نسبت این مدل را پیش‌نهاد دادند [۳]. در این مدل متغیرهای کمکی از طریق یک مدل لوجیتی 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 لوژستیکی (یک مدل برای براورد نسبت بیکاری و یک مدل برای براورد نسبت اشتغال در استان‌ها) براورد کرد. مشکل استفاده از این مدل‌ها این است که ممکن است نسبت‌های براوردشده درون دامنه‌ی [۰,۱] نباشند یا مجموع دو نسبت بیش از یک شود یعنی براورد شاغلان به علاوه‌ی براورد 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 براورد لحاظ نمی‌کنند.

فرض کنید واحدهای آماری هر استان i در رده‌های سنی-جنسیتی j گروه‌بندی می‌شوند و به شرط معلوم بودن p_{ij}ها، مقدارهای متغیر پاسخ y_{ijl} (به ازای $l = ۱, \dots, N_{ij}$ که N_{ij} اندازه‌ی جامعه‌ی استان i و رده‌ی سنی-جنسیتی j است) در خانه‌ی (i,j) جدول متقاطع، متغیرهای برنولی مستقل با احتمال مشترک p_{ij} باشند.

$y_{ijl۱}$ و $y_{ijl۲}$ به‌ترتیب متغیر کمکی مربوط به وضعیت بیکاری و اشتغال فرد و $p_{ij۱}$ و $p_{ij۲}$ به‌ترتیب احتمال‌های بیکاری و اشتغال در رده‌ی سنی-جنسیتی j (= ۱, … , ۶) و استان i (= ۱, … , ۳۰) باشند. دو مدل رگرسیون لوژستیکی (برای شاغلان و برای بیکاران) به‌صورت زیر برازش داده شدند:

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

$$\text{Logit}(p_{ijk}) = \log \frac{p_{ijk}}{۱-p_{ijk}} = X_{ij}\beta_k + v_i,$$

$$i = ۱, \dots, ۳۰, \quad j = ۱, \dots, ۶, \quad k = ۱, ۲ \qquad (۱)$$

که در آن X_{ij} ماتریس معلوم از متغیرهای کمکی (سطح سواد و رده‌های سنی-جنسیتی که در بخش ۳ معرفی خواهند شد)، β_k بردار ضریب‌های رگرسیونی و $v_i \sim N\left(۰, \sigma_v^۲\right)$ اثرهای تصادفی ناحیه‌ای مربوط به استان iام هستند.

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 به‌عنوان رده‌ی پایه در نظر گرفته شد.

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

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

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

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

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

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

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

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

که در آن $\hat{p}^L_{ij۱}$ و $\hat{p}^L_{ij۲}$ به‌ترتیب براورد احتمال‌های بیکاری و اشتغال با استفاده از مدل رگرسیون لوژستیکی و $\hat{y}^r_{ij۱}$ و $\hat{y}^r_{ij۲}$ براورد مجموع بیکاران و شاغلان نانمونه‌ای در رده‌ی سنی-جنسیتی jام و استان iام هستند. اگر N_{ij} و n_{ij} به‌ترتیب اندازه‌های جامعه‌ای و نمونه‌ای در رده‌ی سنی-جنسیتی j و استان iام معلوم باشند، آن‌گاه $n^r_{ij} = N_{ij} - n_{ij}$ اندازه‌ی نانمونه‌ای در این رده‌ها است.

مجموع شاغلان و بیکاران در هر استان را می‌توان با استفاده از عبارت

$$\hat{\delta}^L_{ik} = \sum_{j=۱}^{۶} \left(y_{ijk} + y^r_{ijk}\right), \qquad i = ۱, \ldots, ۳۰, \quad k = ۱, ۲$$

براورد کرد، که در آن $y_{ij۱}$و $y_{ij۲}$ به‌ترتیب مجموع بیکاران و شاغلان نمونه‌ای و $\hat{y}^r_{ij۱}$ و $\hat{y}^r_{ij۲}$ مجموع بیکاران و شاغلان نانمونه‌ای براوردشده در رده‌ی سنی-جنسیتی jام و استان iام هستند.

با استفاده از براورد مجموع بیکاران و شاغلان به دست آمده تحت مدل رگرسیون لوژستیکی، براوردهای نرخ بیکاری در استان‌ها با استفاده از رابطه‌ی

$$\hat{ur}^L_i = ۱۰۰ \times \frac{\hat{\delta}^L_{i۱}}{\hat{\delta}^L_{i۱} + \hat{\delta}^L_{i۲}}, \qquad i = ۱, \ldots, ۳۰.$$

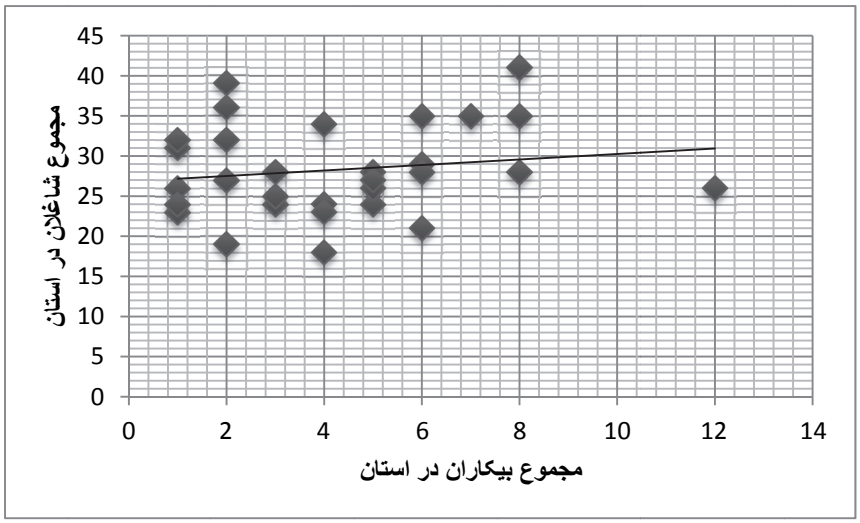
به دست می‌آیند، که در آن $\hat{\delta}^L_{i۱}$ و $\hat{\delta}^L_{i۲}$ به‌ترتیب براورد مجموع بیکاران و شاغلان تحت مدل رگرسیون لوژستیکی در استان i هستند.

۲-۱-۲- مدل آمیخته‌ی لوجیتی چندجمله‌ای

استفاده از مدل آمیخته‌ی خطی یک ابزار عمومی در براوردهای 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 لوژستیکی (یک مدل برای براورد نسبت بیکاری و

یک مدل برای براورد نسبت اشتغال در استان‌ها) براورد 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 به‌صورت 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 براورد مجموع و نسبت‌های بیکاران و شاغلان و همچنین براورد 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 براوردهای مدل‌مبنای هم‌زمان

را برای مجموع بیکاران و شاغلان در کوچک‌ناحیه‌ها فراهم می‌کند. براورد مجموع غیر فعالان اقتصادی با کم کردن از کل جمعیت محاسبه می‌شود.

در کاربرد حاضر در آمارگیری نیروی کار، فرض کنید $y_{ij۱}$، $y_{ij۲}$ و $y_{ij۳}$ به‌ترتیب مجموع بیکاران، شاغلان و غیر فعالان اقتصادی نمونه‌ای در رده‌ی سنی-جنسیتی j و استان i، $n_{ij} = y_{ij۱} + y_{ij۲} + y_{ij۳}$ اندازه‌ی نمونه‌ای رده‌ی سنی-جنسیتی j در استان i و $p_{ij۱}$، $p_{ij۲}$ و $p_{ij۳}$ به‌ترتیب احتمال‌های بیکاری و اشتغال و غیرفعّال اقتصادی بودن در رده‌ی سنی-جنسیتی j و استان i باشند. فرض می‌کنیم بردارهای $(y_{ij۱}$ و $y_{ij۲}$ و $y_{ij۳})$ به شرط معلوم بودن v_i (اثر تصادفی ناحیه‌ای) و n_{ij} مستقل و دارای توزیع چندجمله‌ای با تابع چگالی زیر هستند:

$$f\left(y_{ij۱}, y_{ij۲}\middle| v_i\right) = \frac{n_{ij}!}{y_{ij۱}!, y_{ij۲}!, y_{ij۳}!} {p_{ij۱}}^{y_{ij۱}} {p_{ij۲}}^{y_{ij۲}} {p_{ij۳}}^{y_{ij۳}}. \quad (۲)$$

همچنین فرض می‌کنیم احتمال‌های $p_{ij۱}$ و $p_{ij۲}$ با متغیرهای کمکی سطح سواد و رده‌های سنی-جنسیتی و اثرهای تصادفی ناحیه‌ای دارای رابطه‌ی زیر هستند:

$$\log\left(\frac{p_{ijk}}{p_{ij۳}}\right) = X_{ij}\beta_k + v_i, \quad i = ۱, \dots, ۳۰, \quad j = ۱, \dots, ۶, \quad k = ۱, \dots ۲ \quad (۳)$$

که در آن β_k $(= ۱, ۲)$ 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 $v_i \overset{ind}{\sim} \mathrm{N}(۰, \sigma_v^۲)$.

مدل‌های (۱) و (۲) به مجموع شاغلان و بیکاران نمونه‌ای استان‌ها (به‌صورت توأم) برازش داده می‌شوند. پس از براورد ضریب‌های رگرسیونی و پیش‌گویی اثرهای تصادفی ناحیه‌ای، احتمال‌های بیکاری و اشتغال در رده‌های سنی-جنسیتی و استان از طریق عبارت

$$\hat{p}_{ijk}^{M} = \frac{\exp\left(X_{ij}\hat{\beta}_k + \hat{v}_i\right)}{۱ + \sum_{k=۱}^{۲} \exp\left(X_{ij}\hat{\beta}_k + \hat{v}_i\right)},$$

$$i = ۱, \dots, ۳۰, \quad j = ۱, \dots, ۶, \quad k = ۱, ۲$$

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

با استفاده از احتمال‌های براوردشده‌ی بیکاری و اشتغال، مجموع بیکاران و شاغلان نانمونه‌ای در رده‌ی سنی-جنسیتی jام و استان iام به‌صورت زیر پیش‌گویی می‌شوند:

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

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

$$\hat{\delta}_{ik}^M = \sum_{j=1}^{۶} (y_{ijk} + \hat{y}_{ijk}^r), \qquad i = ۱, \dots, ۳۰, \quad j = ۱, \dots, ۶, \quad k = ۱, ۲$$

که در آن $y_{ij۱}$ و $y_{ij۲}$ به‌ترتیب مجموع بیکاران و شاغلان نمونه‌ای و $\hat{y}_{ij۱}^r$ و $\hat{y}_{ij۲}^r$ به‌ترتیب مقدارهای پیش‌گویی‌شده‌ی بیکاران و شاغلان نانمونه‌ای در رده‌های سنی-جنسیتی jام و استان i هستند.

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

$$\hat{u}r_i^M = ۱۰۰ \times \frac{\hat{\delta}_{i1}^M}{\hat{\delta}_{i۱}^M + \hat{\delta}_{i۲}^M}, \qquad i = ۱, \dots, ۳۰, \quad j = ۱, \dots, ۶, \quad k = ۱, ۲$$

که در آن $\hat{\delta}_{i۱}^M$ و $\hat{\delta}_{i۲}^M$ براورد 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 براورد تغییرات، از روش نمونه‌گیری چرخشی استفاده می‌شود. برای دستیابی به براوردهای 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 به‌دلیل 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 به‌صورت نمونه‌گیری است و اطّلاعات مربوط به خانوار از طریق مصاحبه‌ی حضوری مأمور آمارگیر با مطلع‌ترین عضو خانوار و تکمیل پرسش‌نامه گرداوری می‌شود. اطّلاعات هر یک از اعضای خانوار نیز از طریق مصاحبه‌ی حضوری مأمور آمارگیر با هر یک از اعضای ده ساله و بیش‌تر خانوار گرداوری می‌شود.

برای دستیابی به براوردهای مورد نظر، ابتدا باید داده‌های مربوط به هر یک از خانوارهای نمونه و اعضای آن‌ها وزن‌دهی و سپس از فرمول‌های براورد 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 ارایه‌ی براوردها 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 به‌عنوان جامعه‌ی هدف در نظر گرفته شده است. با استفاده از براورد 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 دقّت براوردها افزایش می‌یابد. بنا بر این با در نظر گرفتن اثر طرح برابر با ۰٫۸ 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.

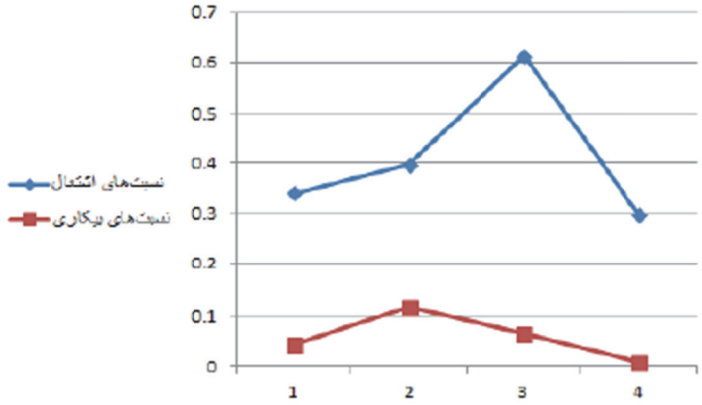
در انتخاب متغیرهای کمکی و به کارگیری آن‌ها در مدل‌های مورد نظر با محدودیت‌های بسیاری مواجه بودیم، زیرا هر چه سطح کوچک‌ناحیه‌ها از استان‌ها کوچک‌تر می‌شوند متغیرهای کمکی محدود می‌شوند. بنا بر این با توجه به پرسش‌نامه متغیرهای کمکی سطح سواد و رده‌های سنی-جنسیتی به‌عنوان متغیرهای کمکی در برازش مدل‌های رگرسیون لوژستیکی و آمیخته‌ی لوجیتی چندجمله‌ای برای براورد 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 کمک

متغییر	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 به‌صورت 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 براورد نسبت‌های بیکاری و اشتغال تأثیرگذار هستند و دقّت براوردها با وام گرفتن از این متغیرهای کمکی افزایش خواهد یافت.

۴- مطالعه‌ی شبیه‌سازی

به‌منظور مقایسه‌ی براوردهای نرخ‌های بیکاری حاصل از مدل‌های رگرسیون لوژستیکی و مدل آمیخته‌ی لوجیتی چندجمله‌ای با هم و با براوردهای مستقیم یک مطالعه‌ی شبیه‌سازی انجام شد.

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

۴-۱- تعیین اندازه‌ی نمونه‌ی اصلی

برای تعیین اندازه‌ی نمونه‌ای 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 به‌صورت تقریبی برابر با ۰٫۱۱۶۸ در نظر گرفتیم.

اندازه‌ی نمونه در سطح کشور در طرح نمونه‌گیری تصادفی ساده به‌صورت

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

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 دقّت براوردها در مقایسه با طرح نمونه‌گیری تصادفی ساده افزایش می‌یابد. لذا برای این که با همین دقّت و با استفاده از طرح نمونه‌گیری طبقه‌ای پارامترها را براورد کنیم اثر طرح را برابر با $\text{deff} = ۰/۸$ در نظر گرفتیم و اندازه‌ی نمونه‌ی طبقه‌ای در سطح کشور را با استفاده از عبارت زیر به دست آوردیم:

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

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).

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

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 قرارگرفته در استان‌ها به این نتیجه می‌رسیم که می‌توان استان‌ها را به‌عنوان کوچک‌ناحیه در نظر گرفت و از فن‌های براورد کوچک‌ناحیه‌ای برای براورد پارامترهای مورد نظر بهره گرفت.

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

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

۱-۴- براوردگرهای مستقیم پارامترهای نیروی کار

فرض کنید n_i و N_i به‌ترتیب اندازه‌ی نمونه‌ای و جامعه‌ای در استان iام و $y_{ij۱}$ و $y_{ij۲}$ به‌ترتیب تعداد بیکاران و شاغلان در نمونه‌های خودگردان رده‌ی سنی-جنسیتی j $(= ۱, \dots, ۶)$ و استان i $(= ۱, \dots, ۳۰)$ باشند، براوردگرهای مستقیم نسبت‌های بیکاری و اشتغال در استان iام در نمونه‌های خودگردان $(= ۱, \dots, ۱۰۰۰)$ bام با استفاده از رابطه‌های زیر به دست آمدند:

$$\hat{p}_{ib۱}^{dir} = \frac{۱}{n_i}\sum_{j=۱}^{n_i} y_{ij۱},$$

$$\hat{p}_{ib۲}^{dir} = \frac{۱}{n_i}\sum_{j=۱}^{n_i} y_{ij۲}, \qquad i = ۱, .., ۳۰$$

براورد مستقیم خودگردان نسبت‌های بیکاری و اشتغال به‌ترتیب عبارت‌اند از:

$$\hat{p}_{iB۱}^{dir} = \frac{۱}{۱۰۰۰}\sum_{b=۱}^{۱۰۰۰} \hat{p}_{ib۱}^{dir},$$

$$\hat{p}_{iB۲}^{dir} = \frac{۱}{۱۰۰۰}\sum_{b=۱}^{۱۰۰۰} \hat{p}_{ib۲}^{dir}, \qquad i = ۱, \dots, ۳۰$$

با استفاده از براورد نسبت‌ها، براوردهای مستقیم مجموع بیکاران و شاغلان به‌صورت

$$\hat{\delta}_{ib۱}^{dir} = N_i \hat{p}_{iB۱}^{dir}, \quad \hat{\delta}_{ib۲}^{dir} = N_i \hat{p}_{iB۲}^{dir}, \qquad i = ۱, \dots, ۳۰$$

حاصل می‌شوند. بنا بر این می‌توان براورد مستقیم نرخ‌های بیکاری در استان‌ها را به‌صورت

$$\hat{u}r_{iB}^{dir} = ۱۰۰ \times \frac{\hat{\delta}_{ib۱}^{dir}}{\hat{\delta}_{ib۱}^{dir} + \hat{\delta}_{ib۲}^{dir}}, \qquad i = ۱, \dots, ۳۰ \qquad (۴)$$

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

$$\hat{MSE}\left(\hat{ur}_i^{dir}\right) = \frac{1}{1000}\sum_{b=1}^{1000}(\hat{ur}_{ib}^{dir} - \hat{ur}_{iB}^{dir})^2, \quad i = 1, \ldots, 30 \qquad (5)$$

براورد کرد که در آن $\hat{ur}_{ib}^{dir}$ براوردگر مستقیم نرخ بیکاری در نمونه‌ی خودگردان bام و $\hat{ur}_{iB}^{dir}$ براورد خودگردان نرخ بیکاری در استان iام هستند.

۴-۲- برازش مدل رگرسیون لوژستیکی به داده‌ها

برای هر نمونه‌ی خودگردان، دو مدل رگرسیون لوژستیکی مجزا به‌صورت مدل (۱) به مجموع بیکاران و شاغلان در رده‌های سنی-جنسیتی و استان‌ها برازش داده شدند. سپس براورد مجموع بیکاران و شاغلان تحت مدل رگرسیون لوژستیکی یعنی $\hat{\delta}_{ib1}^{L}$ و $\hat{\delta}_{ib2}^{L}$ به دست آمدند. نرخ‌های بیکاری در استان‌ها، $\hat{ur}_{ib}^{L}$، با استفاده از رابطه‌ی (۴) براورد شدند(زبرنویس L به مدل رگرسیون لوژستیکی اشاره دارد). همچنین میانگین توان دوم خطا برای نرخ‌های بیکاری تحت مدل‌های رگرسیون لوژستیکی، $\hat{\mathrm{MSE}}(\hat{ur}_i^{L})$ ، با استفاده از رابطه‌ی (۵) تقریب زده شدند.

۴-۳- برازش مدل آمیخته‌ی لوجیتی چندجمله‌ای به داده‌ها

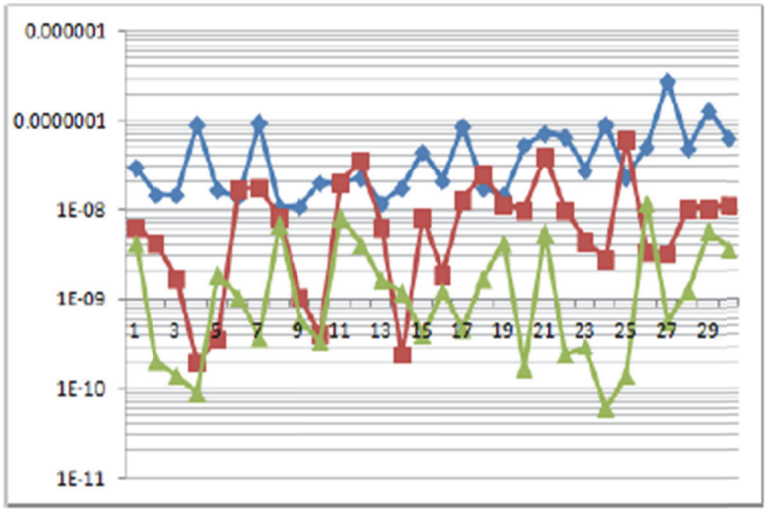
با فرض این که y_{ij1}، y_{ij2} و y_{ij3} به‌ترتیب مجموع بیکاران، شاغلان و غیر فعالان اقتصادی در نمونه‌های خودگردان در رده‌ی سنی-جنسیتی j و استان i، $n_{ij} = y_{ij1} + y_{ij2} + y_{ij3}$ اندازه‌ی نمونه‌ای رده‌ی سنی-جنسیتی j در استان i و p_{ij1}، p_{ij2} و p_{ij3} به‌ترتیب احتمال‌های بیکاری و اشتغال و غیرفعّال اقتصادی بودن در رده‌ی سنی-جنسیتی j در استان i هستند، مدل آمیخته‌ی لوجیتی چندجمله‌ای (۲) و (۳) به نمونه‌های خودگردان برازش داده شد و نرخ‌های بیکاری تحت مدل آمیخته‌ی لوجیتی چندجمله‌ای، $\hat{ur}_{ib}^{M}$، با استفاده از عبارت (۴) براورد شدند (زبرنویس M به مدل لوجیتی آمیخته‌ی چندجمله‌ای اشاره دارد). سپس MSE براوردهای نرخ بیکاری در کوچک‌ناحیه‌ی i، $\hat{\mathrm{MSE}}(\hat{ur}_i^{M})$، با استفاده از رابطه‌ی (۵) تقریب زده شدند.

براورد خودگردان MSEی براوردهای نرخ‌های بیکاری حاصل از دو مدل رگرسیون لوژستیکی و مدل آمیخته‌ی لوجیتی چندجمله‌ای در جدول ۳ نشان داده شده‌اند.

جدول ۳- براورد خودگردان MSEی براوردهای نرخ‌های بیکاری مدل رگرسیون لوژستیکی و مدل آمیخته‌ی لوجیتی چندجمله‌ای و براورد مستقیم

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

شکل ۴ میانگین توان دوم خطای خودگردان براوردهای مستقیم و مدل‌مبنای 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 براوردهای مدل‌مبنا در تمام استان‌ها کوچک‌تر از $\widehat{MSE}$ی براوردهای مستقیم است. همچنین $\widehat{MSE}$ی براوردهای حاصل از مدل آمیخته‌ی لوجیتی چندجمله‌ای در بسیاری از استان‌ها کوچک‌تر از $\widehat{MSE}$ی براوردهای حاصل از مدل رگرسیون لوژستیکی است زیرا این مدل همبستگی بین مجموع بیکاران و شاغلان در استان‌ها را در فرایند براورد منظور کرده است.



شکل ۴- میانگین توان دوم خطای خودگردان برآوردهای مستقیم و مدل‌مبنای نرخ‌های بیکاری در مقیاس لگاریتمی استان‌ها: میانگین توان دوم خطای براوردگرهای خودگردان نرخ بیکاری استان‌ها: مستقیم ♦، تحت مدل رگرسیون لوژستیکی ▲ و تحت مدل آمیخته‌ی لوجیتی چندجمله‌ای ■

۵- بحث و نتیجه‌گیری

در این مقاله پس از معرفی دو مدل رگرسیون لوژستیکی و مدل آمیخته‌ی لوجیتی چندجمله‌ای در یک مطالعه‌ی شبیه‌سازی این دو مدل به داده‌ها برازش داده شدند. یافته‌های مطالعه‌ی شبیه‌سازی نشان می‌دهد که اولاً براوردهای مدل‌مبنا به‌دلیل یاری گرفتن از متغیرهای کمکی در مقایسه با براوردهای مستقیم دقّت بیش‌تری دارند. ثانیاً

براوردهـای نرخ‌هـای بیکـاری تحـت مـدل آمیختـه‌ی لـوجیتی چنـدمتغیره در مقایسـه بـا براوردهای نرخ‌های بیکاری بر اساس مدل رگرسیون لوژستیکی دقیق‌تر هسـتند؛ زیـرا ایـن 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@srtc.ac.ir