برآورد و پیش‌بینی کوچک‌ناحیه‌ای میانگین طول مدت بیکاری در ایران و اثر استان بر آن با استفاده از مدل‌های سه‌سطحی

سید محمدابراهیم حسینی‌نسب[۱] و راضیه احمدلو

بخش آمار، دانشگاه شهید بهشتی

تاریخ دریافت:۹۲/۵/۱۲ تاریخ پذیرش:۹۳/۳/۲۰

چکیده: نیاز به برآوردهای کوچک‌ناحیه‌ای از طریق داده‌های آمارگیری از مدت‌ها قبل احساس شده است. اهمیت استفاده از چنین برآوردهایی در این است که اندازه‌ی نمونه در بعضی از نواحی کوچک ممکن است کم یا حتی صفر شود طوری که به‌موجب آن فرایندهای آماری در داخل هر کوچک‌ناحیه به برآوردهای مستقیم با دقت قابل‌قبول منجر نشود. اگر داده‌های درون ناحیه‌های کوچک دارای ساختار سلسله‌مراتبی باشند، باید از مدل‌های چندسطحی برای ایجاد برآوردهایی با دقت قابل‌قبول استفاده نمود. در این مقاله داده‌های آمارگیری نیروی کار مرکز آمار ایران در سال ۱۳۸۸ برای هشت استان مختلف تحلیل شده است. در این مطالعه متغیر پاسخ که طول مدت بیکاری افراد بر حسب روز است، بر حسب متغیرهای کمکی سن، جنسیت، بستگی با سرپرست خانوار، وضع تحصیلات، وضع تأهل، سطح تحصیلات و تجربه‌ی کار قبلی مدل‌بندی شده است. چون ساختار داده‌ها به‌صورت سلسله‌مراتبی است برای مدل‌بندی متغیر پاسخ از مدل‌های سه‌سطحی استفاده کردیم. در این مدل‌بندی، متغیرهای وضع تحصیلات و وضع تأهل معنادار نبودند. همچنین اثرهای ثابت و مؤلفه‌های واریانس را با روش ماکسیمم درستنمایی مقید برآورد کرده و اثرهای تصادفی را به شیوه‌ی BLUP پیشگویی کرده‌ایم و سپس نتایج اصلی گزارش و تفسیر شده‌اند. در پایان نیز با استفاده از این مدل، برآورد کوچک-ناحیه‌ای میانگین طول مدت بیکاری برای هشت استان مختلف پیش‌گویی شده است.

واژه‌های کلیدی: مدل چندسطحی، برآورد کوچک‌ناحیه‌ای، طول مدت بیکاری.

رده‌بندی ریاضی:۶۲J۰۷، ۶۲J۹۹.

۱– آدرس الکترونیکی نویسنده مسئول مقاله: سید محمدابراهیم حسینی‌نسب m_hosseininasab@sbu.ac.ir

۱- مقدمه

دیر زمانی است که آمارگیری‌های نمونه‌ای برای فراهم نمودن برآوردهایی از مجموع‌ها، میانگین‌ها و دیگر پارامترها به طور گسترده به کار گرفته می‌شوند. این آمارگیری‌ها به طور گسترده برای فراهم کردن برآوردهایی برای زیرجامعه‌ها، ناحیه‌ها یا حوزه‌ها به‌کار می‌روند. علاوه بر مشکلات عملی، اجرای آمارگیری‌های نمونه‌ای که برآوردهای قابل‌اعتماد را در سطح‌هایی بهتر از قلمرو کشوری، استانی و غیره فراهم می‌کنند، نیازمند افزایش اندازه‌ی نمونه است که به‌طور کلی پرهزینه هستند. بنابراین روش‌های برآورد کوچک‌ناحیه‌ای برای غلبه بر مشکل کوچک بودن اندازه‌ی نمونه مورد استفاده قرار می‌گیرند .

در حالت کلی روش‌های برآورد کوچک‌ناحیه‌ای را می‌توان این‌گونه تعریف کرد: روش‌های برآورد کوچک‌ناحیه‌ای، روش‌های ایجاد برآوردهای قابل‌اعتماد مناسب برای نواحی جغرافیایی است که در آن‌ها استفاده از روش‌های برآورد مستقیم دقت مورد نیاز را ایجاد نمی‌کند. استفاده از برآورد مستقیم، به معنی استفاده از روش‌های برآورد نمونه‌ای طرح پایه‌ی کلاسیک است که از واحدهای نمونه‌ی موجود در همان ناحیه استفاده می‌کنند. هر ناحیه‌ای که نمونه‌ی آن به اندازه‌ی کافی بزرگ نباشد تا برآوردهای مستقیم قابل‌اعتماد و با دقت کافی تولید کند به عنوان کوچک‌ناحیه تلقی می‌شود. در برخی کاربردها، بسیاری از حوزه‌های مورد نظر ممکن است حتی دارای اندازه‌ی نمونه‌ی صفر باشند (مانند نوعی بیماری خاص در حوزه‌ی بعضی شهرستان‌ها). بنابراین از روش‌های برآورد کوچک‌ناحیه‌ای می‌توان برای تهیه‌ی آمارهای قابل‌اعتماد برای ناحیه‌های کم‌نمونه یا بدون نمونه استفاده نمود.

به این علت که برآوردگرهای مستقیم دقت مورد نیاز را برای تولید برآوردهای کوچک‌ناحیه‌ای فراهم نمی‌کنند، اغلب از برآوردگرهای غیرمستقیم استفاده می‌شود. برآوردگرهای غیرمستقیم علاوه بر استفاده از واحدهای نمونه یا مقدار متغیر مورد نظر همان ناحیه، از مقدارهای متغیر مورد نظر ناحیه‌ها یا دوره‌های زمانی مرتبط دیگر و همچنین اطلاعات کمکی مرتبط با متغیر مورد نظر مانند اطلاعات آخرین سرشماری، اطلاعات ثبتی جاری و ... استفاده می‌کنند. بنابراین در حالت کلی می‌توان گفت در دسترس بودن داده‌های کمکی خوب و تعیین مدل‌های پیونددهنده‌ی مناسب برای ساختن برآوردگرهای غیرمستقیم از عامل‌های اساسی روی آوردن به برآوردهای کوچک‌ناحیه‌ای هستند. می‌توان برای اطلاعات بیشتر راجع به برآوردهای کوچک‌ناحیه‌ای به [۱-۴] مراجعه کرد.

در روش‌های تحلیل واریانس و تحلیل رگرسیون معمولاً بعضی فرض‌های پایه‌ای مانند استقلال مشاهدات مطرح است. گاهی اوقات این فرض برای داده‌های مورد مطالعه صادق نیست و در

نتیجه به‌کارگیری این مدل‌ها مناسب نخواهد بود. مدل‌های چندسطحی یک رویکرد مدل‌سازی جدید است که به ما امکان نادیده گرفتن فرض استقلال و مدل‌سازی داده‌هایی با ساختار سلسله مراتبی را می‌دهد. این مدل‌ها برای تحلیل داده‌هایی که در آن گروه‌ها به‌صورت تصادفی هستند، مناسب است. در تحقیقاتی همچون [۴-۸]، روش‌های تجزیه و تحلیل داده‌هایی با ساختار سلسله مراتبی، گسترش داده شدند.

در بعضی موارد، زمانی که ساختار داده‌های درون ناحیه‌های کوچک به صورت سلسله‌مراتبی باشد باید از مدل‌های چندسطحی برای برآورد کوچک‌ناحیه‌ای استفاده نمود. به عنوان مثالی از ساختار سه‌سطحی، می‌توان به افراد درون شهرها و شهرهای درون استان‌ها اشاره نمود که در آن افراد در سطح یک، شهرها در سطح دو و استان‌ها در سطح سه قرار دارند و افراد، شهرها و استان‌ها به تصادف انتخاب می‌شوند. مدل‌های چندسطحی از داده‌های کمکی نیز در سطوح متفاوت استفاده می‌کنند و با استفاده از اثرهای تصادفی در مدل، تغییرات بین ناحیه‌ها را در نظر می‌گیرند.

در برآورد اثرهای ثابت و پیشگویی اثرهای تصادفی در مدل‌های چندسطحی برای برآورد کوچک‌ناحیه‌ای، بهترین برآوردگر نااریب خطی (BLUE[1]) و بهترین پیشگوی نااریب خطی (BLUP[2]) مورد استفاده قرار می‌گیرند. BLUP برای یک ترکیب خطی از اثرهای ثابت و تصادفی اولین بار توسط هندرسن در [۹] معرفی شد. بعدها در [۱۰]، با جاگذاری پارامترهای برآورد شده در بهترین پیشگوی پیشنهادشده در [۹] یک پیشگوی نااریب دومرحله‌ای پیشنهاد شد که به آن بهترین پیشگوی نااریب خطی تجربی (EBLUP[3]) می‌گویند. بعد از آن در سال ۲۰۰۹ و در [۱۱] فرم تعمیم‌یافته‌ی BLUP ارائه‌شده در [۹] معرفی شد.

در این مقاله با استفاده از مدل سه‌سطحی برای برآورد کوچک‌ناحیه‌ای، طول مدت بیکاری افراد مدل‌بندی می‌شود. مدلی که در این مقاله مورد استفاده قرار گرفته است با استفاده از داده‌های نیروی کار مرکز آمار ایران که در بهار سال ۱۳۸۸ از استان‌های مختلف جمع‌آوری شده‌اند، می‌باشد. جزئیات مدل سه‌سطحی و روش‌های برآورد پارامترهای آن در بخش ۲ بیان می‌شود. در بخش ۳ مدل‌های سه‌سطحی برای برآورد کوچک‌ناحیه‌ای و برآوردگر میانگین کوچک‌ناحیه‌ای مورد بررسی قرار می‌گیرد. در بخش ۴ توضیحاتی در مورد نحوه‌ی جمع‌آوری داده‌ها، آمارگیری نیروی کار و متغیرهای مورداستفاده ارائه می‌شود. در بخش ۵ نیز نتایج استفاده از این مدل‌ها آورده شده است و در نهایت در بخش ۶ نتیجه‌گیری ارائه می‌شود.

---

1 -Best Linear Unbiased Estimator
2 -Best Linear Unbiased Predictor
3 -Empirical Best Linear Unbiased Predictor

۲- مدل‌های سه‌سطحی و برآورد پارامترهای آن

بسیاری از داده‌ها مانند داده‌های جمع‌آوری شده در حوزه‌های علوم اجتماعی، انسانی، زیستی و پزشکی دارای ساختار سلسله‌مراتبی چندسطحی هستند. به عنوان مثالی برای داده‌های سه‌سطحی می‌توان به داده‌هایی اشاره کرد که در آن دانش‌آموزان از کلاس‌ها و کلاس‌هایی از مدرسه‌ها انتخاب شده‌اند آن چنان که در آن دانش‌آموزان در سطح یک، کلاس‌ها در سطح دو و مدرسه‌ها در سطح سه قرار دارند. در این حالت ابتدا تعدادی مدرسه به تصادف انتخاب، سپس تعدادی کلاس از داخل این مدارس به تصادف برگزیده‌شده، در نهایت تعدادی دانش‌آموز از داخل این کلاس‌ها به‌تصادف انتخاب خواهد شد.

در یک مدل سه‌سطحی فرض می‌شود که سطح یک آن شامل k فرد، سطح دو آن شامل i گروه و سطح سه آن دارای j گروه است. این مدل سه‌سطحی با متغیر پاسخ y و یک متغیر تبیینی x برای فرد iام درون گروه jام است، به‌صورت زیر تعریف می‌شود:

$$y_{ijk} = \beta_{0ij} + \beta_{1ij} x_{ijk} + \varepsilon_{ijk}, \; k = 1,\dots,n_{ij}, \; i = 1,\dots,I, \; j = 1,\dots,J, \qquad (1)$$

که در آن y_{ijk} و x_{ijk} به ترتیب متغیر پاسخ و متغیر تبیینی برای فرد kام در گروه iام درون گروه jام هستند. ε_{ijk}ها عبارت‌های خطای تصادفی هستند که فرض می‌شود به‌صورت نرمال با میانگین صفر و واریانس σ_e^2 توزیع شده‌اند. β_{0ij} و β_{1ij} به ترتیب عرض از مبدأ و شیب رگرسیونی برای گروه iام درون گروه jام هستند. پارامترهای β_{0ij} و β_{1ij} در میان گروه‌ها تابعی از یک متغیر تبیینی دیگر w_{ij} در سطح گروه (سطح دو) هستند. به‌طور دقیق‌تر β_{0ij} و β_{1ij} برای هر گروه به‌صورت زیر تعریف می‌شوند:

$$\begin{aligned} \beta_{0ij} &= \alpha_{00j} + \alpha_{01j} w_{ij} + u_{0ij}, \\ \beta_{1ij} &= \alpha_{10j} + \alpha_{11j} w_{ij} + u_{1ij}. \end{aligned} \qquad (2)$$

در عبارت‌های تصادفی β_{0ij} و β_{1ij} عبارت‌های خطای u_{0ij} و u_{1ij} تصادفی فرض می‌شوند. به‌زبانی دقیق‌تر فرض می‌شود که عبارت‌های تصادفی u_{0ij} و u_{1ij} به‌صورت نرمال با میانگین صفر، واریانس σ_{u0}^2 و σ_{u1}^2 و کوواریانس σ_{u01} توزیع شده‌اند. در معادله‌ی (۲) پارامترهای α_{00j} و α_{10j} عرض از مبدأ و α_{01j} و α_{11j} شیب رگرسیونی برای گروه jام هستند و در میان گروه‌ها تابعی از متغیر تبیینی دیگر (z_j) در سطح سه هستند که به‌صورت زیر تعریف می‌شوند:

$$\begin{aligned} \alpha_{00j} &= \gamma_{000} + \gamma_{001} z_j + \upsilon_{00j}, \\ \alpha_{01j} &= \gamma_{010} + \gamma_{011} z_j + \upsilon_{01j}, \\ \alpha_{10j} &= \gamma_{100} + \gamma_{101} z_j + \upsilon_{10j}, \\ \alpha_{11j} &= \gamma_{110} + \gamma_{111} z_j + \upsilon_{11j}. \end{aligned} \qquad (۳)$$

عبارت‌های خطای υ_{00j}، υ_{01j}، υ_{10j} و υ_{11j} تصادفی دارای توزیع نرمال با میانگین صفر، واریانس به‌ترتیب $\sigma^2_{\upsilon_{00}}$، $\sigma^2_{\upsilon_{01}}$، $\sigma^2_{\upsilon_{10}}$ و $\sigma^2_{\upsilon_{11}}$ و دوبه‌دو دارای کوواریانس هستند. با جاگذاری معادلات (۳) در معادلات (۲) و سپس با جاگذاری معادله‌ی به‌دست آمده در معادله‌ی (۱)، معادله‌ی کلی برای این مورد از مدل سه‌سطحی، به‌صورت زیر به‌دست می‌آید:

$$\begin{aligned} y_{ijk} = {} & \gamma_{000} + \gamma_{001} z_j + \upsilon_{00j} + \gamma_{010} w_{ij} + \gamma_{011} z_j w_{ij} + \upsilon_{01j} w_{ij} + \gamma_{100} x_{ijk} \\ & + \gamma_{101} z_j x_{ijk} + \upsilon_{10j} x_{ijk} + \gamma_{110} w_{ij} x_{ijk} + \gamma_{111} z_j w_{ij} x_{ijk} + \varepsilon_{ijk}. \end{aligned} \qquad (۴)$$

لازم به ذکر است که در حالت کلی ممکن است بیش‌تر از یک متغیر تبیینی در سطح اول و همچنین بیش‌تر از یک متغیر تبیینی در سطح دوم وجود داشته باشد.

روشی که به طور معمول در تحلیل رگرسیون چندسطحی برای برآورد پارامترها مورد استفاده قرار می‌گیرد روش ماکسیمم درستنمایی (ML) است. به‌طور متداول دو نوع روش ML مختلف برای تحلیل رگرسیون چندسطحی مورد استفاده قرار می‌گیرد. روش اول ماکسیمم درستنمایی کامل (FML[1]) نامیده می‌شود که در این روش ضرایب رگرسیونی و مؤلفه‌های واریانس در تابع درستنمایی حضور دارند. روش دیگر روش ماکسیمم درستنمایی مقید (RML[2]) است که در این روش فقط مؤلفه‌های واریانس در تابع درستنمایی قرار می‌گیرند. تفاوت بین این دو روش آن است که روش FML برآوردهای ضرایب رگرسیونی را برای برآورد مؤلفه‌های واریانس به عنوان مقداری معلوم در نظر می‌گیرد اما روش RML آن‌ها را به عنوان برآوردهایی که مقدار نامعلومی دارند تلقی می‌کند [۱۲ و ۱۳]. از این رو RML واقع‌بینانه‌تر است اما در عمل اختلاف‌های بین دو روش خیلی زیاد نیست [۷].

1 -Full Maximum Likelihood
2 -Restricted Maximum Likelihood

۳- برآورد کوچک‌ناحیه‌ای با استفاده از مدل‌های سه‌سطحی

مدل کلی سه‌سطحی

$$\begin{aligned} y_j &= X_j\beta_j + \varepsilon_j \\ \beta_j &= V_j\lambda_j + u_j, \quad j = ۱,\ldots,J, \\ \lambda_j &= W_j\gamma + \alpha_j, \end{aligned} \qquad (۵)$$

را در نظر بگیریم که در آن y_j برداری از متغیر پاسخ برای واحدهای نمونه در j امین کوچک-ناحیه، X_j ماتریس متغیرهای تبیینی در سطح واحد آماری نمونه، V_j و W_j ماتریس طرح متغیرهای کوچک‌ناحیه به‌ترتیب در سطح دو و سه، γ برداری شامل ضرایب ثابت، u_j و α_j بردارهای شامل اثرهای تصادفی به ترتیب در سطح دو و سه برای j امین کوچک‌ناحیه هستند. به علاوه فرض‌های زیر درباره‌ی توزیع بردارهای تصادفی در نظر گرفته می‌شوند:

۱) u_jها بین کوچک‌ناحیه‌ها، مستقل از یکدیگر هستند و دارای یک توزیع توأم مشترک با $E(u_j) = ۰$ و $Var(u_j) = D$ درون هر کوچک‌ناحیه هستند. همچنین α_jها نیز بین کوچک‌ناحیه‌ها مستقل از یکدیگر و دارای یک توزیع مشترک با $E(\alpha_j) = ۰$ و $Var(\alpha_j) = C$ درون هر کوچک‌ناحیه هستند.

۲) ε_jها از یکدیگر مستقل هستند، $E(\varepsilon_j) = ۰$ و $Var(\varepsilon_j) = \sigma_e^۲ I$.

۳) ε_jها، u_jها و α_jها از یکدیگر مستقل هستند.

با جاگذاری معادله سطح سه در سطح دو و همچنین با جاگذاری معادله سطح دو در سطح یک، مدل (۵) را می‌توان به صورت زیر نوشت:

$$y_j = X_jV_jW_j\gamma + X_jV_j\alpha_j + X_ju_j + \varepsilon_j. \qquad (۶)$$

با تغییر متغیر $Z_j = V_jW_j$ و $\upsilon_j = V_j\alpha_j + u_j$ مدل (۶) را می‌توان به صورت زیر نوشت:

$$y_j = X_jZ_j\gamma + X_j\upsilon_j + \varepsilon_j,\ j = ۱,\ldots,J. \qquad (۷)$$

مجموعه‌ی j معادله در رابطه‌ی (۷) می‌تواند به طور مختصر و فشرده به فرم ماتریسی زیر نوشته شود:

$$y = XZ\gamma + X\upsilon + \varepsilon \qquad (۸)$$

که در آن $y=(y'_1,y'_2,\ldots,y'_J)'$ ، $\varepsilon=(\varepsilon'_1,\varepsilon'_2,\ldots,\varepsilon'_J)'$ ، $Z=(Z'_1,Z'_2,\ldots,Z'_J)'$ ، و $\upsilon=(\upsilon'_1,\upsilon'_2,\ldots,\upsilon'_J)'$

$$X=\begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_J \end{bmatrix}.$$

یکی از روش‌های محاسبه‌ی برآورد γ و پیشگویی υ استفاده از معادلات مدل آمیخته است [۱۴]. در این روش از لگاریتم تابع چگالی توأم y و υ نسبت به γ و υ مشتق گرفته و با مساوی صفر قرار دادن معادلات و حل همزمان آن‌ها BLUE(γ) و BLUP(υ) به‌دست می‌آیند.

۳-۱- برآورد میانگین کوچک‌ناحیه‌ای

مدل (۷) را در نظر بگیریم؛ اگر فرض کنیم که اندازه‌ی جامعه N_j برای j امین کوچک‌ناحیه، به‌اندازه‌ی کافی بزرگ است، با توجه به این که

$$\mu_j=\bar{y}_j=\bar{X}'_j Z_j\gamma+\bar{X}'_j\upsilon_j+\bar{\varepsilon}_j,$$

و $\bar{\varepsilon}_j\approx 0$ که در آن $\bar{\varepsilon}_j$ میانگین N_j خطای ε_{ij} است، می‌توان میانگین j امین کوچک‌ناحیه را به‌صورت زیر نوشت:

$$\mu_j=\bar{X}'_j Z_j\gamma+\bar{X}'_j\upsilon_j, \tag{۹}$$

که در آن $\bar{X}_j$ بردار میانگین متغیرهای تبیینی با طول $p+1$ برای j امین کوچک‌ناحیه است.

یک برآوردگر برای μ_j را می‌توان با جایگزین کردن برآوردگرهای ML، REML یا RIGLS[1] برای γ و پیشگویی υ_j به ترتیب در جملات معادله‌ی (۹) به‌دست آورد. لازم به ذکر است که برآوردگر RIGLS تحت فرض نرمال بودن هم‌ارز با برآوردگر REML است [۱۵]. برآوردگر μ_j زیر به عنوان بهترین برآوردگر نااریب خطی تجربی (EBLUE) شناخته می‌شود [۱۶]:

$$\hat{\mu}_j=\bar{X}'_j Z_j\hat{\gamma}+\bar{X}'_j\hat{\upsilon}_j, \tag{۱۰}$$

1 -Restricted Iterative Generalized Least Squares

که در آن $\hat{\gamma}=(\sum_{j=1}^{J} Z_j' X_j' \hat{V}_j^{-1} X_j Z_j)^{-1} \sum_{j=1}^{J} Z_j' X_j' \hat{V}_j^{-1} y$. پیشگوی اثر تصادفی برای j - امین کوچک‌ناحیه برابر است با $\hat{\upsilon}_j=\hat{D}X_j'\hat{V}_j^{-1}(y_j - X_j Z_j \hat{\gamma})$ که در آن $\hat{V}_j^{-1}=\hat{\sigma}^{-2} I - \hat{\sigma}^{-4} X_j \hat{D} \hat{G}_j^{-1} X_j'$ و $\hat{G}_j^{-1}=(I+\hat{\sigma}^{-2} X_j' X_j \hat{D})^{-1}$. در [۱۷ و ۱۸] یک مدل عرض از مبدأ تصادفی برای برآوردهای کوچک‌ناحیه‌ای پیشنهاد شد. این مدل عبارت است از:

$$y_j = X_j \gamma + \upsilon_{j0} 1_j + \varepsilon_j ,$$

که در آن $1_j=(1,1,\ldots,1)'$ بردارهایی با طول n_j برای واحدهای نمونه‌گیری شده در j امین، $j=1,\ldots,J$ ، کوچک‌ناحیه است. X_j یک ماتریس $n_j \times (p+1)$ از متغیرهای تبیینی است. بردار γ یک برداری با طول $(p+1)$ از پارامترهای ثابت رگرسیون و υ_{j0} یک اثر تصادفی اسکالر برای هر کوچک‌ناحیه است به‌طوری که $E(\upsilon_{j0})=0$ ، $Var(\upsilon_{j0})=\sigma_0^2$ و برای، $Cov(\upsilon_{j0},\upsilon_{j'0})=0$. همچنین فرض می‌شود $E(\varepsilon_j)=0$ ، $Var(\varepsilon_j)=\sigma_e^2 I$ و ε_jها مستقل از هم و نیز مستقل از υ_{j0}ها هستند. در این حالت، بهترین برآوردگر نااریب خطی تجربی به‌صورت زیر است:

$$\hat{\mu}_{j(RI)} = \bar{X}_j' \hat{\gamma} + \hat{\upsilon}_{j0} .$$

اندیس RI برای نشان دادن مدل عرض از مبدأ تصادفی، زمانی که فقط عرض از مبدأ هر کوچک‌ناحیه تصادفی و سایر مؤلفه‌های γ ثابت هستند، به‌کاربرده می‌شود. برای جزئیات بیشتر در این زمینه می‌توان به [۱۶] مراجعه نمود.

۴- داده‌های آمارگیری نیروی کار

مسئله‌ی بیکاری در هر جامعه‌ای از اهمیت ویژه‌ای برخوردار است. بیکاری مسئله‌ای کلان در اقتصاد است که به طور مستقیم و به‌شدت بر افراد جامعه تأثیر می‌گذارد. برای بسیاری از افراد ازدست دادن شغل به معنی کاهش سطح زندگی و آزردگی خاطر است و به این دلیل بیکاری یکی از متداول‌ترین مباحث سیاسی می‌باشد، به‌گونه‌ای که سیاستمداران همیشه ادعا می‌کنند سیاست‌های پیشنهادی آنان می‌تواند مشاغل جدید به وجود آورد.

در بررسی پدیده‌ی بیکاری طول مدت آن از اهمیت بسیاری برخوردار است. مطالعات انجام‌شده نشان می‌دهند که بیکاری در اکثر جوامع وجود داشته و در زمان‌های مختلف طول مدت آن متفاوت است به‌علاوه بیکاری در گروه‌های مختلف جمعیتی یا نواحی مختلف یک کشور یکسان نیست. بنابراین با در نظر گرفتن اثر بیکاری بر گروه‌ها و نواحی مختلف جوامع، مدت بیکاری

شاخص مهمی در مطالعه‌ی آن است. بیکاری کوتاه مدت ممکن است به دلیل زمان‌بر بودن فرایند جستجوی شغل و در نتیجه عملکرد طبیعی بازار کار به‌وجود آید و در نتیجه غیرقابل اجتناب است. از سوی دیگر بیکاری بلندمدت و مزمن نشانه‌ی وجود عدم تطابق بین مشاغل و مهارت‌های موجود در بازار کار یا کمبود تقاضای نیروی کار از سوی بنگاه‌های اقتصادی است.

با توجه به مطالب ذکرشده، برای پاسخگویی به نیاز روزافزون متولیان برنامه‌ریزی اقتصادی کشور، لازم است که برآوردهایی برای طول مدت بیکاری در سطوح کوچک‌تر با استفاده از روش‌های کوچک‌ناحیه‌ای ارائه شود. در کنار این فعالیت‌ها، در این مقاله سعی می‌شود با استفاده از داده‌های طرح آمارگیری نیروی کار، برآوردهایی برای طول مدت بیکاری به شیوه‌ی مدل‌بندی سه‌سطحی ارائه شود. برای این منظور از داده‌های آمارگیری نیروی کار مرکز آمار ایران که در بهار سال ۱۳۸۸ جمع‌آوری شده است، استفاده می‌شود. آمارگیری نیروی کار، یک آمارگیری نمونه‌ای فصلی و یکی از مهم‌ترین آمارگیری‌های نمونه‌ای در ایران است. طراحی این آمارگیری بر اساس آخرین توصیه‌های سازمان بین‌المللی کار (ILO) انجام‌شده است و برای اولین بار در سال ۱۳۸۴ با ۴۴۵۶۸ خانوار نمونه در هر فصل اجرا شده تا با استفاده از آن بتوان آمارهای جاری نیروی کار کشور و همچنین تغییرات مرتبط را در دوره‌های زمانی کوتاه به‌دست آورد. این آمارگیری به‌منظور تأمین نیازهای آماری سیاست‌گذاران، تصمیم‌گیران و محققان در بخش‌های اقتصادی و اجتماعی برای ارزیابی سیاست‌های قبلی دولت و برنامه‌ریزی برای تصمیمات آینده اجرا می‌شود.

جامعه‌ی مورد مطالعه، افراد بیکار کشور در بهار ۱۳۸۸ است و مدت بیکاری آن‌ها تا زمان آمارگیری مورد بررسی قرار گرفت. ابتدا از میان ۳۰ استان کشور، به تصادف هشت استان انتخاب که استان‌های نمونه‌گیری شده شامل آذربایجان شرقی، خراسان رضوی، سمنان، فارس، قم، کردستان، گیلان و هرمزگان هستند. سپس از درون هر کدام از این استان‌ها، تعدادی منطقه شهری و روستایی به تصادف انتخاب، و از درون هر کدام افراد بیکار نمونه‌گیری شده‌اند، بنابراین ماهیت این داده‌ها به صورت سه‌سطحی است.

۴-۱- معرفی متغیرهای مدل

در این پژوهش متغیر وابسته طول مدت بیکاری افراد است. همچنین متغیرهای کمکی که در این بررسی مورد استفاده قرار می‌گیرند، به صورت زیر هستند:

- سن
- جنسیت :۱- مرد و ۲- زن

- بستگی با سرپرست خانوار: ۱- سرپرست و ۲- بقیه‌ی اعضای خانوار
- وضع تحصیلات: ۱- در حال تحصیل نبودن و ۲- در حال تحصیل بودن
- وضع تأهل: ۱- ازدواج‌کرده و ۲- طلاق گرفته، بیوه یا ازدواج‌نکرده
- سطح تحصیلات: ۱- زیر دیپلم، ۲-دیپلم یا پیش‌دانشگاهی، ۳- فوق‌دیپلم و لیسانس و ۴- فوق‌لیسانس و دکترا
- تجربه‌ی کار قبلی: ۱- تجربه داشتن و ۲- تجربه نداشتن

در جدول‌های ۱ و ۲ خلاصه‌ی آماری متغیرهای فوق برای استان‌های مختلف به تفکیک نقاط شهری و روستایی ارائه شده است (ردیف‌های دوم و بعد از آن، تعداد را نشان می‌دهند).

جدول ۱: خلاصه‌ی آماری متغیرهای مدل برای نقاط شهری هشت استان مختلف

استان / متغیر	آذربایجان	فارس	قم	گیلان	هرمزگان	خراسان	کردستان	سمنان
میانگین سن	۲۸/۲۹	۲۷/۴۰	۲۶/۳۲	۲۸/۶۵	۲۴/۷۰	۲۷/۹۰	۲۸/۹۹	۲۷/۷۴
مرد	۴۶	۱۱۶	۷۵	۷۵	۲۰	۷۶	۹۹	۴۵
زن	۳۲	۶۶	۲۱	۶۶	۱۷	۳۶	۱۹	۲۱
سرپرست	۱۹	۳۳	۲۰	۱۸	۵	۳۲	۳۵	۱۳
بقیه اعضای خانوار	۵۹	۱۴۹	۷۶	۱۲۳	۳۲	۸۰	۸۳	۵۳
در حال تحصیل بودن	۱۲	۱۶۸	۸۸	۱۳۴	۳۶	۱۰۵	۱۱۸	۶۳
در حال تحصیل نبودن	۶۶	۱۴	۸	۷	۱	۷	۰	۳
ازدواج‌کرده	۲۷	۷۱	۳۲	۴۵	۸	۵۲	۴۵	۲۲
ازدواج‌نکرده	۵۱	۱۱۱	۶۴	۹۶	۲۹	۶۰	۷۳	۴۴
زیر دیپلم	۲۷	۷۰	۳۵	۳۳	۱۳	۴۷	۵۴	۲۶
دیپلم یا پیش‌دانشگاهی	۲۰	۶۰	۳۷	۶۴	۱۸	۲۷	۲۸	۱۷
فوق‌دیپلم و لیسانس	۲۸	۴۶	۲۱	۴۳	۶	۲۶	۲۱	۲۲
فوق‌لیسانس و دکترا	۳	۶	۳	۱	۰	۱۲	۱۵	۱
تجربه داشتن	۴۷	۹۳	۵۸	۷۳	۱۸	۷۵	۸۸	۴۳
تجربه نداشتن	۳۱	۸۹	۳۸	۶۸	۱۹	۳۷	۳۰	۲۳

جدول ۲: خلاصه‌ی آماری متغیرهای مدل برای نقاط روستایی هشت استان مختلف

استان / متغیر	آذربایجان	فارس	قم	گیلان	هرمزگان	خراسان	کردستان	سمنان
میانگین سن	۲۷/۱۸	۲۷/۱۴	۲۴/۸۵	۲۷/۶۱	۲۶/۷۳	۲۵/۳۳	۲۶/۱۸	۲۹/۲۷
مرد	۱۸	۷۸	۲۱	۳۴	۱۶	۵۷	۱۱	۱۶
زن	۹	۲۱	۰	۱۸	۳	۱۱	۰	۲
سرپرست	۷	۳۰	۴	۸	۴	۲۰	۱	۴
بقیه اعضای خانوار	۲۰	۶۹	۱۷	۴۴	۱۵	۴۸	۱۰	۱۴
در حال تحصیل بودن	۲۳	۹۱	۲۱	۵۱	۱۹	۶۸	۱۱	۱۸
در حال تحصیل نبودن	۴	۸	۰	۱	۰	۰	۰	۰
ازدواج‌کرده	۱۳	۴۳	۴	۲۰	۵	۲۹	۴	۵
ازدواج‌نکرده	۱۴	۵۶	۱۷	۳۲	۱۴	۳۹	۷	۱۳
زیر دیپلم	۱۶	۶۲	۱۱	۲۴	۱۱	۴۷	۴	۱۰
دیپلم یا پیش‌دانشگاهی	۶	۱۵	۸	۱۵	۳	۱۰	۵	۳
فوق‌دیپلم و لیسانس	۵	۱۴	۰	۱۱	۲	۵	۲	۵
فوق‌لیسانس و دکترا	۰	۸	۲	۲	۳	۶	۰	۰
تجربه داشتن	۱۳	۵۵	۱۴	۳۲	۱۰	۵۳	۹	۱۴
تجربه نداشتن	۱۴	۴۴	۷	۲۰	۹	۱۵	۲	۴

۵- تحلیل داده‌های آمارگیری نیروی کار

در این بخش از مدل‌های سه‌سطحی برای برآوردهای کوچک‌ناحیه‌ای مربوط به داده‌های نیروی کار استفاده می‌شود. در ابتدای این بخش برآورد مستقیم میانگین طول مدت بیکاری در هشت استان ارائه می‌شود. در ادامه نیز یک مدل سه‌سطحی را به داده‌ها برازش داده و به تحلیل عددی داده‌ها می‌پردازیم و در نهایت با استفاده از مدل، برآورد کوچک‌ناحیه‌ای طول مدت بیکاری در هشت استان مختلف را به‌دست می‌آوریم.

۵-۱- برآورد مستقیم میانگین طول مدت بیکاری در استان‌ها

برآوردهای مستقیم تنها از داده‌های نمونه‌ای ناحیه‌ها استفاده می‌کنند به این صورت که هر ناحیه تنها از مقدار متغیرهای مورد نظر مربوط به واحدهایی را که در آن ناحیه قرار دارند استفاده می‌کند. در این قسمت با استفاده از داده‌های طول مدت بیکاری هر استان برآورد میانگین را به‌دست آورده‌ایم. برآوردهای مستقیم میانگین طول مدت بیکاری برای هشت استان مختلف در جدول ۳ گزارش شده است.

جدول ۳: برآورد مستقیم میانگین طول مدت بیکاری (بر حسب روز) برای هشت استان مختلف

استان	برآورد مستقیم میانگین طول مدت بیکاری
آذربایجان شرقی	۶۶۰
فارس	۷۵۳
قم	۶۴۴
گیلان	۹۰۱
هرمزگان	۵۵۲
خراسان رضوی	۳۲۶
کردستان	۴۰۱
سمنان	۴۸۷

مقادیر جدول یانگر این مطلب هستند که میانگین طول مدت بیکاری افراد در استان‌های گیلان و فارس زیاد است و در استان‌های خراسان رضوی و کردستان میانگین طول مدت بیکاری افراد کمتر از بقیه استان‌ها است.

۵-۲- مدل‌بندی سه‌سطحی داده‌ها

به دلیل ماهیت سه‌سطحی بودن داده‌ها در این بخش از مدل‌های سه‌سطحی برای مدل‌بندی طول مدت بیکاری بر حسب متغیرهای کمکی مربوطه استفاده می‌کنیم. فرض کنیم متغیر mb_{ijk} مدت بیکاری را در استان iام، نقاط شهری و روستایی jام و تکرار kام نشان دهد. مدل سه‌سطحی که شامل اثرهای تصادفی استان و نقاط شهری و روستایی داخل استان‌ها است، به صورت زیر است:

$$\begin{aligned} mb_{ijk} &= \beta_0 + \beta_1 age_{ijk} + \beta_2 sex_{ijk} + \ldots + \beta_9 \,\text{experience}_{ijk} \qquad i = 1,\ldots,8,\ j = 1,2 \\ &\quad + \upsilon_i + \upsilon_{i,j} + \varepsilon_{ijk}, \qquad k = 1,\ldots,n_{ij}, \end{aligned} \qquad (11)$$

که در آن age، sex و experience به‌ترتیب نشان‌دهنده‌ی متغیرهای سن، جنسیت و تجربه‌داشتن است و همچنین فرض می‌شود که $\upsilon_i \sim N(0,\sigma_1^2)$، $\upsilon_{i,j} \sim N(0,\sigma_2^2)$ و $\varepsilon_{ijk} \sim N(0,\sigma_e^2)$. اثر تصادفی استان‌ها، υ_iها، مستقل از هم و اثر تصادفی نقاط شهری و روستایی درون استان‌ها، $\upsilon_{i,j}$ها، برای i و j های مختلف مستقل از هم هستند. به‌علاوه، فرض می‌شود υ_iها و $\upsilon_{i,j}$ها نیز مستقل از هم هستند. خطاهای درون گروهی، ε_{ijk}ها نیز برای i، j و k های مختلف مستقل از هم و نیز مستقل از اثرهای تصادفی دیگر هستند.

جدول۴: معیارهای نیکویی برازش مدل

logLik	BIC	AIC
-۲۳۶۵/۳۱۳	۴۸۰۸/۰۲۳	۴۷۵۲/۶۲۵

جدول ۵: برآورد انحراف معیار اثرهای تصادفی مدل

انحراف معیار اثرهای تصادفی	استان	نقاط شهری/روستایی	مانده‌ها
برآورد	۱/۲۸۱	۱/۷۷۹	۰/۸۸۲

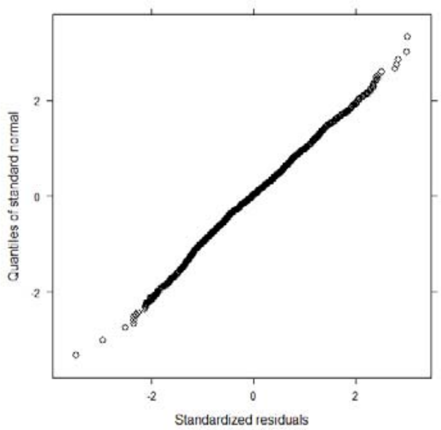
با برازش مدل (۱۱) به داده‌های مورد نظر و با بررسی کفایت مدل با رسم نمودار احتمال نرمال مانده‌ها و آزمون کولموگروف-اسمیرنوف ملاحظه می‌شود که فرض نرمال بودن مانده‌ها و ثبات واریانس برقرار نیست. با تبدیل ۰/۳ داده‌ها ملاحظه می‌شود که فرض نرمال بودن مانده‌ها و ثبات واریانس برقرار می‌گردد ولی در این مدل برازش داده‌شده باملاحظه‌ی p–مقدار مشخص می‌شود که اثرهای وضع تحصیلات و وضع تأهل معنی‌دار نیستند بنابراین با خارج کردن این اثرها مدل دیگری بدون وجود این اثرها به داده‌ها برازش می‌دهیم، که نتایج حاصل از برازش این مدل به صورت زیر است:

جدول۶: برآورد اثرهای ثابت مدل

اثرهای ثابت	برآورد	St.error	اثرهای ثابت	برآورد	St.error
عرض از مبدأ	۳/۴۹۷	۰/۴۲۶	سطح تحصیلات		
سن	۰/۰۹۵	۰/۰۰۹	مبنا: فوق‌لیسانس و دکترا	–	–
			زیر دیپلم	۰/۷۷۰	۰/۲۸۵
جنسیت			دیپلم یا پیش‌دانشگاهی	۱/۳۶۷	۰/۲۹۷
مبنا: زن	–	–	فوق‌دیپلم و لیسانس	۱/۱۰۶	۰/۳۰۴
مرد	۰/۵۸۱	۰/۱۵۱			
بستگی با سرپرست خانوار			تجربه‌ی کار قبلی		
مبنا: سایر اعضای خانوار	–	–	مبنا: تجربه نداشتن	–	–
سرپرست خانوار	۱/۶۳۴−	۰/۱۹۹	تجربه داشتن	۲/۰۱۱−	۰/۱۳۶

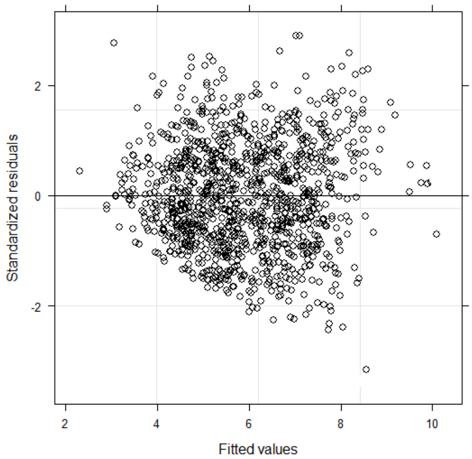
۵-۳- بررسی کفایت مدل

ابتدا فرض نرمال بودن داده‌ها را با رسم نمودار احتمال نرمال مانده‌ها بررسی می‌نماییم. نمودار احتمال نرمال شکل ۱ نشان می‌دهد که مانده‌ها دارای توزیع نرمال هستند. همچنین p–مقدار آزمون کولموگروف-اسمیرنوف برای مانده‌های این مدل برابر با ۰/۷۹۸۴ است که بیانگر نرمال بودن توزیع این مانده‌ها است.

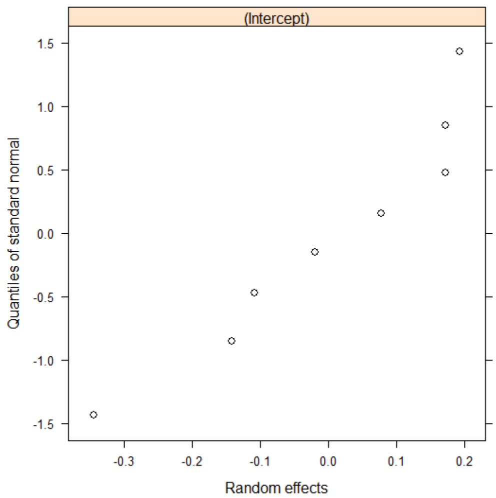


شکل ۱: نمودار احتمال نرمال برای مانده‌ها

فرض بعدی بررسی ثبات واریانس است که برای بررسی این موضوع باید نمودار مانده‌ها در برابر مقادیر پیش‌بینی‌شده را رسم نماییم. شکل ۲ نشان می‌دهد که مانده‌ها حول صفر پراکنده‌شده‌اند و هیچ الگویی در این نمودار وجود ندارد و بنابراین هیچ گواهی برای انحراف از فرض ثبات واریانس وجود ندارد.



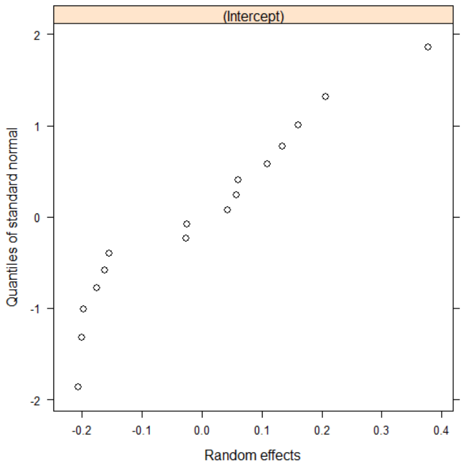
شکل ۲: نمودار پراکنش مانده‌ها در برابر مقادیر برازش داده‌شده در مدل



شکل ۳: نمودار احتمال نرمال اثرهای تصادفی برآورد شده‌ی استان‌ها برای مدل

در بررسی کفایت برازش یک مدل چندسطحی، باید فرض نرمال بودن اثرهای تصادفی برآورد شده در هر یک از سطوح گروه‌بندی نیز بررسی شود. نمودار احتمال نرمال اثرهای تصادفی برآورد شده‌ی استان‌ها برای مدل در شکل ۳ نشان داده‌شده است. به‌دلیل آن‌که تنها هشت سطح برای اثر تصادفی استان داریم، تشخیص انحراف از فرض نرمال بودن مشکل است اما به نظر می‌آید که فرض نرمال بودن اثرهای تصادفی برآورد شده‌ی استان‌ها برقرار است. p–مقدار آزمون کولموگروف-اسمیرنوف برابر با ۰/۱۲۹ است که بیانگر نرمال بودن توزیع اثرهای تصادفی برآورد شده‌ی استان‌ها است.

نمودار احتمال نرمال اثرهای تصادفی نقاط شه ری و روستایی به صورت آشیانه‌ای درون استان‌ها در شکل ۴ نشان داده‌شده است.



شکل ۴: نمودار احتمال نرمال اثرهای تصادفی برآورد شده‌ی نقاط شهری و روستایی درون استان

در این نمودار نیز گواهی بر انحراف از فرض نرمال بودن وجود ندارد. همچنین p–مقدار آزمون کولموگروف-اسمیرنوف برای اثرهای تصادفی برآورد شده‌ی نقاط شهری و روستایی درون استان‌ها برابر با ۰/۲۰۰ است که نشان‌دهنده‌ی نرمال بودن توزیع آن‌ها است.

۵-۴- تفسیر نتایج مدل

با توجه به نتایج به‌دست آمده از مدل، انحراف معیار نقاط شهری و روستایی درون استان‌ها ۱/۷۷۹ است و انحراف معیار درون گروه‌ها ۰/۸۸۲ است که نشان می‌دهد تفاوت بین نقاط

شهری و روستایی درون استان‌ها بیشتر از تفاوت درون گروه‌ها است. همچنین انحراف معیار بین استان‌ها ۱/۲۸۱ است که بیانگر این مطلب است که بین استان‌های مختلف تفاوت وجود دارد.

با استفاده از داده‌های جدول ۶ مشخص است که در صورت ثابت بودن سایر شرایط، به ازای افزایش یک سال سن طول مدت بیکاری تبدیل‌یافته به نسبت ۰/۰۹۵ افزایش می‌یابد. همچنین مشخص می‌شود که سرپرست خانوار طول مدت بیکاری کم‌تری نسبت به سایر اعضای خانوار دارد و همچنین افرادی که تجربه‌ی کار قبلی دارند نسبت به افرادی که تجربه ندارند دارای طول مدت بیکاری کم‌تری هستند. اثر سطح تحصیلات نیز نشان می‌دهد افراد در سطح دیپلم یا پیش‌دانشگاهی طول مدت بیکاری تبدیل‌یافته بیشتری دارند.

۵-۵- ارزیابی اثرهای تصادفی مدل

در این بخش پیشگویی اثرهای تصادفی به شیوه‌ی BLUP را در هر یک از سطوح اثرها ارائه می‌دهیم. پیشگویی‌های مربوط به اثرهای تصادفی در سطح استان‌ها در جدول ۷ آمده است.

جدول ۷: پیشگویی اثرهای تصادفی در سطح استان‌ها

استان	آذربایجان	فارس	قم	خراسان	هرمزگان	گیلان	سمنان	کردستان
پیشگویی	۰/۱۹۶	۰/۲۰۹	۰/۰۸۲	۰/۳۵۹-	۰/۱۴۹-	۰/۱۷۴	۰/۲۲۰-	۰/۱۲۱-

همان‌طور که از جدول ۷ ملاحظه می‌شود استان‌های خراسان رضوی، هرمزگان و کردستان مدت بیکاری کم‌تری دارند و کم‌ترین طول بیکاری را استان خراسان رضوی دارد. استان‌های آذربایجان شرقی، فارس و گیلان مدت بیکاری بیشتری دارند و بیش‌ترین طول مدت بیکاری را استان فارس دارد. در جدول ۸ پیشگویی اثرهای تصادفی در سطح نقاط شهری و روستایی داخل استان‌ها ارائه شده است:

جدول ۸: پیشگویی اثرهای تصادفی در سطح نقاط شهری و روستایی درون استان

استان	آذربایجان	فارس	قم	خراسان	هرمزگان	گیلان	سمنان	کردستان
شهری	۰/۰۱۶-	۰/۰۳۸	۰/۰۲۴-	۰/۱۹۰-	۰/۰۶۲	۰/۳۷۹	۰/۱۵۶-	۰/۱۷۸-
روستایی	۰/۲۰۹	۰/۱۶۰	۰/۱۰۵	۰/۱۶۵-	۰/۲۱۰-	۰/۲۰۶-	۰/۱۳۲	۰/۰۵۷

با توجه به نتایج جدول ۸ می‌توان گفت که اثر نقاط شهری و روستایی در استان‌های مختلف متفاوت است. نقاط شهری استان‌های آذربایجان شرقی، قم، خراسان رضوی، سمنان و کردستان مدت بیکاری کم‌تری دارند و کم‌ترین مدت بیکاری را نقاط شهری استان خراسان رضوی دارد. نقاط شهری استان‌های فارس، هرمزگان و گیلان مدت بیکاری بیش‌تری دارند و بیش‌ترین مدت بیکاری را استان گیلان دارد.

نقاط روستایی استان‌های خراسان رضوی، هرمزگان و گیلان مدت بیکاری کم‌تری دارند و کم-ترین مدت بیکاری را نقاط روستایی استان هرمزگان دارد. نقاط روستایی استان‌های آذربایجان شرقی، فارس، قم، سمنان و کردستان مدت بیکاری بیش‌تری دارند و بیش‌ترین مدت بیکاری را نقاط روستایی استان آذربایجان شرقی دارد.

۵-۶- برآورد میانگین طول مدت بیکاری

در این قسمت برآوردهای کوچک‌ناحیه‌ای میانگین طول مدت بیکاری را که با استفاده از مدل (۱۱) برای استان‌های مختلف به‌دست آمده است، ارائه می‌دهیم. برآوردهای جدول ۹ نشان می‌دهد که میانگین طول مدت بیکاری افراد در استان‌های گیلان و فارس زیاد است و استان‌های خراسان رضوی و کردستان دارای میانگین طول مدت بیکاری کم‌تری هستند.

جدول ۹: برآورد کوچک‌ناحیه‌ای میانگین طول مدت بیکاری (بر حسب روز) برای هشت استان با استفاده از مدل

استان	برآورد میانگین طول مدت بیکاری
آذربایجان شرقی	۵۰۱
فارس	۵۲۷
قم	۴۴۳
گیلان	۶۴۸
هرمزگان	۴۰۰
خراسان رضوی	۲۴۴
کردستان	۳۱۹
سمنان	۳۹۸

۶- بحث و نتیجه‌گیری

در این مقاله، از مدل سه‌سطحی برای برآورد کوچک‌ناحیه‌ای استفاده شد. همچنین از نرم‌افزارهای موجود (مانند R) برای برآورد پارامترهای مدل و پیشگویی اثرهای تصادفی در داده‌های بیکاری نیروی کار (بهار ۱۳۸۸) مرکز آمار ایران استفاده شده است. در این داده‌ها متغیر پاسخ طول مدت بیکاری افراد است و با استفاده از مدل‌بندی داده‌ها در این مقاله به این واقعیت دست یافتیم که افرادی که سرپرست خانوار هستند طول مدت بیکاری کم‌تری نسبت به سایر اعضای خانوار دارند و همچنین افرادی که تجربه‌ی کار قبلی دارند نسبت به افرادی که تجربه ندارند دارای طول مدت بیکاری کم‌تری هستند. اثر سطح تحصیلات نیز نشان می‌دهد افراد در سطح دیپلم یا پیش‌دانشگاهی طول مدت بیکاری تبدیل‌یافته بیش‌تری دارند. در انتهای این مقاله نیز برآوردهای کوچک‌ناحیه‌ای میانگین طول مدت بیکاری بر حسب روز برای هشت استان مختلف را با استفاده از مدل به‌دست آوردیم که در آن برآورد میانگین طول مدت بیکاری افراد در استان‌های گیلان، فارس، آذربایجان شرقی و قم زیاد است و استان‌های خراسان رضوی، کردستان، هرمزگان و سمنان دارای میانگین طول مدت بیکاری کم‌تر هستند .

مراجع

[1] Rao, J. (2001). Small Area Estimation with Applications to Agriculture, In Proceedings of the Second *Conference on Agricultural and Environmental Statistical Applications*, ISTAT, Rome, Italy.

[2] Pefefferman, D. (2002). Small Area Estimation-New Developments and Direction, *International Statistical Review,***70**, 125-143.

[3] Rao, J. (2003). *Small Area Estimation*, New York: Wiley.

[4] Aitkin, M. and Longford, N. (1986). Statistical Modeling Issues in School Effectiveness Studies, *Journal of the Royal Statistical Society, (series A)*, **149**, 1-43.

[5] Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares, *Biometrika,* **73**, 43-56

[6] Raudenbush, S. and Bryk, A. (1986). A Hierarchical Model for Studying School Effects, *Sociology of Education,* **59**, 1-17.

[7] Hox, J. (2002). *Multilevel Analysis: Techniques and Applications*, Lawrence Erlbaum Associates, New Jersey.

[8] Afshartous, D. and Leeuw, J. (2005), Prediction in Multilevel Models, *Journal of the Educational and Behavioral Statistics*, **30**, 109--139.

[9] Henderson. C. (1975). Best Linear Unbiased Estimation and Prediction Under aSelection Model, *Biometrics*, **31**,423-447.

[10] Kackar, R. and Harville, D. (1981). Unbiasedness of Two-stage Estimation and Prediction Procedures for Mixed Linear Models,*Communications in Statistics, Series A*, **10**, 1249-1261.

[11] Zadlo, T. (2009). On MSE of EBLUP, *Statistical Papers*, **50**, 101--118.

[12] Bryke, A. and Raudenbush, S. (1992). *Hierarchical Linear Models: Applications and data analysis methods,* California: Sage Publication, Newbury Park.

[13] Goldstein, H. (1995). *Multilevel Statistical Models*, London: Edward Arnold, New York: Halsted.

[14] Henderson. C. (1959). Estimation of Environmental and Genetic Trends from Records Subjects to Culling, *Biometrics*, **15**, 192-218.

[15] Goldstein, H. (1989). Restricted Unbiased Iterative Generalized Least Squares Estimation, *Biometrika,* **76**, 622-623.

[16] Moura, F., and Holt, D. (1999). Small Area Estimation Using Multilevel Models, *Survey Methodology*, **25**, 73-80.

[17] Battese, G. and Fuller, W. (1981). Prediction of County Crop Areas Using Survey and Satellite Data, *Proceedings of the Section on Survey Research Methods*, American Statistical Association. 500-505.

[18] Battese, G., Harter, R., and Fuller, W. (1988). An Error Component Model for Prediction of County Crop Areas Using Survey and Satelite Date,*Journal of the American Statistical Association***, 83**, 28--36.

Small Area Estimation and Prediction of Unemployment Duration Mean In Iran and the Effect of the Provinces On It Using Three-level Models

Seyed Mohammad Ebrahim Hosseininasab and Razieh Ahmadloo

Department of Statistics, Shahid Beheshti University, Tehran, Iran

Abstract

The need for small area estimates via survey data has long been recognized. The importance of using such estimates is that the sample size in some of small areas may be small or even zero so that direct estimates cannot be made with acceptable precision in each small area. If data within small areas have hierarchical structure, multilevel models should be used to make estimates with acceptable precision. In this paper, the labor force survey data of 8 different Iranian provinces, collected by statistical center of Iran in 1388 (Iranian solar year), have been analyzed. In this study, response variable that is unemployment duration of the persons in the survey, is modeled according to auxiliary variables such as age, sex, connection with householder, education situation, marital status, educational level and previous labor experience. Since data structure is hierarchical, we used three-level models for modeling the response variable. In this modeling, the education situation and marital status variables are not significant. The fixed effects and variance components are also estimated by restricted maximum likelihood method and the random effects are predicted by the BLUP method and then the main results are reported and interpreted. Finally using this model, the small area estimations of unemployment duration for 8 different provinces are predicted.

Keywords: Multilevel model, Small area estimation, Unemployment duration, Restricted maximum likelihood.

Mathematics Subject Classification (2010): 62J07, 62J99.