

فصلنامه پژوهش های اقتصادی ایران / شماره ۱۱ / تابستان ۸۱

پیشنهادی برای برآورد نسبت بیکاری در سطح استان های کشور

* حمیدرضا نواب پور*

** علیرضا خوشگویان فرد**

*** محمد طاهری منزه***

چکیده

معمولاً دفاتر ملی آمار کشورها، به دلیل محدودیت هایی از قبیل بودجه و زمان سعی می کنند طرح های آمارگیری خود را در سطوح جغرافیایی بزرگ مثلاً ، در سطح ملی، طراحی و اندازه نمونه ها را بهینه کنند. این امر، سبب می شود که پژوهشگران اقتصادی، برنامه ریزان، و تصمیم گیرندگان که نیاز به اطلاعات آماری در سطوح کوچکتر جغرافیایی و یا به تفکیک صفاتی خاص دارند از دستیابی به اطلاعات مورد نیاز محروم بمانند. کاربران این نوع اطلاعات با انتقال نیازهای آماری خود به واحدهای تولید کننده آمار موجب شده اند که این واحدها، طرح های آمارگیری خود را در سطوح کوچکتر جغرافیایی - مثلاً در سطح استان - طراحی و اندازه نمونه ها را بهینه کنند. این امر، افزون بر تحمیل هزینه به بودجه عمومی کشورها و طولانی شدن فرایند تهیه و توزیع آمار، موجب می شود که خطاهای غیر نمونه گیری سهم بیشتری در مراحل تهیه و توزیع آمارها داشته باشند. در این مقاله، سعی می شود با معرفی چند روش برآورد ناحیه کوچک^۱ (SAE) روشی مناسب برای برآورد نسبت بیکاری در سطح استان های کشور پیشنهاد شود.

واژه های کلیدی: برآوردگر بیز تجربی، برآوردگر غیرمستقیم، برآورد حافظ ساختار، برآوردگر ترکیبی، طبقات پسین، ناحیه کوچک.

Email: h.navvabpour@src.ac.ir

* استادیار دانشگاه علامه طباطبائی

** محقق مرکز تحقیقات صدا و سیما جمهوری اسلامی ایران

*** کارشناس مرکز آمار ایران

1. Small Area Estimation

۱. مقدمه

بیشتر آمارگیری‌های نمونه‌ای که به طور مستمر اجرا می‌شوند، به منظور ارایه برآورد در مورد مناطق از پیش تعیین شده، مثلاً در سطح ملی، طراحی می‌شوند. این در حالی است که اکثر پژوهشگران اقتصادی، برنامه‌ریزان، و تصمیم‌گیرندگان نیازمند استفاده از آمارهای قابل اعتماد در سطوح کوچکتر جغرافیایی و یا به تفکیک صفتی خاص مثلاً شهری و روستایی، هستند. به منظور صرفه‌جویی در زمان و هزینه‌ها، غالباً از این آمارگیری‌ها به طور مستقیم برای ارایه برآوردهای مربوط به زیر جامعه‌های (نواحی کوچک) مناطق مذکور بهره‌گیری می‌شود، اگرچه تعداد نمونه‌ها در ناحیه کوچک به دلیل کوچک بودن نقش معرف بودن خود را ایفا نمی‌کنند. به همین دلیل از آمارشناسان و جمعیت‌شناسان انتظار می‌رود که برآوردهای قابل اعتمادی را بدون در اختیار داشتن داده‌های لازم در مورد زیر ناحیه‌ها به کاربران ارایه کنند. از این رو، آنها برای تأمین این نیاز روش‌هایی را با عنوان روش‌های برآورد ناحیه کوچک ابداع کردند. وجه مشترک کلیه این روش‌ها که آنها را مطلوبتر از برآوردگر مستقیم ساخته، بهره‌گیری از داده‌های مکمل علاوه بر داده‌های نمونه است.

هدف اصلی از کاربرد روش‌های برآورد ناحیه کوچک، ارایه برآوردهای قابل اعتماد برای زیر مجموعه‌های یک جمعیت است، در صورتی که اندازه نمونه برای کل جمعیت تعیین شده باشد. مثلاً، مرکز آمار ایران به علت نیاز فزاینده به برآورد نسبت بیکاری در سطح استان‌های کشور، چند سالی است که برای برآورد نسبت بیکاری در سطح استان‌ها، سالیانه اقدام به اجرای آمارگیری‌های نمونه‌ای در سطح استان‌ها می‌کند، در صورتی که اقتصاددانان به صورت ماهانه و یا حداقل به صورت فصلی به این آمارها نیاز دارند، و این نیاز به دلیل محدودیت‌های بودجه‌ای، زمانی، حرفه‌ای، اجرایی، پردازش و اطلاع‌رسانی قابل تأمین نیست، مگر با به کارگیری روش‌های برآورد ناحیه کوچک.

باید خاطر نشان ساخت که برآوردهای مستقیم این زیر مجموعه‌ها (نواحی کوچک)، نمی‌توانند به علت کوچکی اندازه نمونه (خطای بزرگ نمونه‌گیری)، برآوردهای قابل اعتمادی به دست دهند. چندین برآوردگر غیرمستقیم در طی تقریباً چهل سال پیشنهاد شده است. برای تعریف مسئله و بررسی روش‌های برآورد ناحیه کوچک، خواننده می‌تواند به پرسل و کیش^۱ (۱۹۷۹، ۱۹۸۰)، گوش و راتو^۲ (۱۹۹۴)، شیبیل^۳ (۱۹۹۵)، مارکر^۴ (۱۹۹۹)، و پفرمان^۱ (۱۹۹۹) مراجعه کند.

1. Purcell & Kish
2. Ghosh & Rao
3. Schaible
4. Marker

در ایران، مطالعه روش‌های برآورد ناحیه کوچک در سازمان‌های آماری و سایر سازمان‌های تولید کننده آمار نادیده گرفته شده است. در سال‌های اخیر این موضوع، زمینه‌های جالبی را برای پژوهش فراهم آورده است. در این مقاله، سه روش برآورد ناحیه کوچک ترکیبی، مرکب، و بیز تجربی به منظور برآورد نسبت بیکاری در سطح استان‌های کشور- در حالی که اندازه نمونه در سطح کل کشور بهینه شده است- در نظر گرفته می‌شود.

ابتدا، از راه برآورد حافظ ساختار، برآوردگر ترکیبی مورد بررسی قرار می‌گیرد. استفاده از برآوردگر ترکیبی به علت سادگی محاسبات و سازگاری با طرح‌های رایج نمونه‌گیری بسیار متداول است. این برآوردگر، بر "ترکیب" اطلاعات پیشین ناحیه‌های کوچک و برآوردگرهای نمونه بخش‌های دودو ناسازگار و جامع از جامعه‌ای که نواحی کوچک را در برمی‌گیرد، متکی است. به همین دلیل است که این برآوردگر به عنوان برآوردگر ترکیبی شناخته شده است. مطالعات تجربی نشان داده‌است که این برآوردگر اریب ولی نسبت به برآوردگر مستقیم دارای واریانس کوچکتری است. استفاده از اصطلاح برآورد ترکیبی را به مرکز ملی آمارهای بهداشتی آمریکا منتسب می‌کنند که به نظر می‌رسد اولین استفاده مستند از این برآوردگر در آنجا صورت گرفته است. سپس، برآوردگر مرکب به صورت مجموع موزون برآوردهای مستقیم و ترکیبی بررسی و در نهایت، روش بیز تجربی به طور خلاصه ذکر می‌شود.

در تحلیل کاربردی، یک طرح نمونه‌گیری را که برای برآورد نسبت بیکاری در سطح کشور طراحی شده است مورد توجه قرار می‌دهیم. سپس، با بهره‌گیری از داده‌های به دست آمده از ۲۸ استان کشور برآوردهای خود را ازایه می‌کنیم و عملکرد مربوط به هر یک از آنها را مورد مقایسه قرار می‌دهیم. مبنای مقایسه این برآوردها، برآوردی است که با بهینه کردن اندازه نمونه در سطح استان به دست می‌آید.

۲. مروری بر برآوردها

۲-۱. برآورد حافظ ساختار^۲

این برآوردگر اساساً، مقادیر اولیه پارامترهای مربوط به نواحی کوچک را که از سرشماری قبل به دست آمده است، روز آمد می‌کند که این امر، از روی داده‌های حاصل از آمارگیری‌های نمونه‌ای بزرگ

(یا سایر منابع روز آمد) انجام می شود. در اینجا روشی که متناسب به پرسل و کیش (۱۹۸۰) است و چارچوب آن در واقع برای داده‌های شمارشی طراحی شده است، در نظر گرفته می شود. فرض کنید افزون بر متغیر مورد نظر که به عنوان برآوردنی^۱ شناخته می شود، یک یا چند متغیر کمکی دیگر نیز که با این متغیر در حد بالایی مرتبط هستند، در اختیار باشند. برای به کار بردن روش مطلوب نیاز به دو مجموعه اطلاعات زیر است که رابطه‌های بین این برآوردنی و متغیرهای کمکی را سازمان دهد.

۱. ساختار پیوند

این ساختار، رابطه‌ها را در سطح ناحیه کوچک مورد بررسی، سازمان می‌دهد. نظر به اینکه داده‌های جاری برای نواحی کوچک موجود نیست، این ساختار، از روی داده‌های سرشماری ایجاد می‌شود و در قالب یک جدول پیشابندی سه بُعدی با فراوانی خانه‌های N_{ijh} که در آن اندیس‌ها به ترتیب به i امین ناحیه کوچک ($i = 1, 2, \dots, I$)، j امین گروه متغیر کمکی ($j = 1, 2, \dots, J$) و h امین گروه برآوردنی ($h = 1, 2, \dots, H$) مربوط می‌شود، به نمایش در می‌آید.

۲. ساختار انتساب

این ساختار رابطه‌ها را در سطح جامعه (یا زیر جامعه‌ای که ناحیه‌های کوچک را شامل می‌شود) سازماندهی می‌کند. این ساختار، با یک جدول پیشابندی دوبعدی با فراوانی خانه‌های m_{jh} (که در آن اندیس‌ها به صورت قبل تعریف می‌شوند)، نشان داده می‌شود. توجه داشته باشید که می‌توان m_{jh} ها را به عنوان حاشیه‌های بهنگام شده جدول قبلی همراه با $N_{\bullet jh}$ ها محسوب کرد ($N_{\bullet jh}$ ها شمارش‌های جدول هستند که روی اندیس j جمع بسته شده‌اند)، با این دلیل که $N_{\bullet jh}$ ها از روی داده‌های قدیمی و m_{jh} ها از روی داده‌های جاری به دست می‌آیند.

هدف، برآورد (یا اصلاح) N_{ijh} هایی است که با m_{jh} ها (حاشیه‌های بهنگام شده) سازگار باشند و روابط موجود در ساختار پیوند را حفظ کنند. چنانچه، فراوانی‌های خانه‌ای برآورد شده را با $N = \{N_{ijh}\}$ نشان دهیم، هدف نهایی روی حاشیه‌هایی از X است که مربوط به نواحی کوچک هستند یعنی $X_{i,h}$. حال، فرض کنید که $N = \{N_{ijh}\}$ و $M = \{m_{jh}\}$ ، مشخص باشند. برای

برآورد، می‌توان از روش کمترین توان‌های دوم موزون برای مینیمم کردن مجموع توان‌های دوم موزون تفاوت‌های بین N_{ijh} و X_{ijh} نسبت به قیدهای حاشیه‌ای که با ساختار انتساب معین می‌شوند، استفاده کرد به این معنی که تابع لاگرانژ زیر را مینیمم کرد:

$$F(X_{ijh}, \lambda) = \sum_{i,j,h} W_{ijh}^{-1} (X_{ijh} - N_{ijh})^2 - \lambda \left(\sum_i X_{ijh} - m_{jh} \right) \quad (1)$$

که در آن W_{ijh} ها، وزن‌های از پیش تعیین شده‌اند و λ ضریب لاگرانژ است. چنانچه مشتقات $F(X_{ijh}, \lambda)$ نسبت به X_{ijh} و λ برابر صفر قرار داده شود، رابطه زیر به دست می‌آید:

$$X_{ijh} = W_{ijh} W_{.jh}^{-1} (m_{jh} - N_{.jh}) + N_{ijh} \quad (2)$$

با انتخاب $W_{ijh} = N_{ijh}$ فراوانی خانه‌ای برآورد شده به صورت زیر است:

$$X_{ijh} = \frac{N_{ijh}}{N_{.jh}} m_{jh} \quad (3)$$

با جمع بستن (۳) روی بعد متغیر کمکی، برآورد برای h امین گروه برآوردنی در i امین ناحیه کوچک از رابطه زیر به دست می‌آید:

$$X_{i,h} = \sum_j \frac{N_{ijh}}{N_{.jh}} m_{jh} \quad (4)$$

و در نهایت، برآورد حافظ ساختار برای نسبت مورد نظر به شرح زیر حاصل می‌شود:

$$\hat{P}_{i,h}^s = \frac{X_{i,h}}{N_{i..}} = \sum_j \frac{N_{ijh}}{N_{.jh} N_{i..}} m_{jh} \quad (5)$$

۲-۲. برآوردگر ترکیبی^۱

رابطه (۴) را به تعریف سنتی برآوردگر ترکیبی که به اختصار در مقدمه شرح داده شد، مرتبط می‌کنیم. برای برآورد، در درجه اول، جامعه‌ای که بر نواحی کوچک احاطه دارد به چند طبقه تحت عنوان طبقه‌های پسین براساس متغیری که کاملاً با متغیر مورد نظر مرتبط است، افزایش می‌دهد. چنانچه از طبقه‌های پسین برآوردهای نارایب مستقیم در اختیار باشد، برآورد ترکیبی پارامتری که مورد بررسی است برابر با مجموع موزون برآوردهای طبقه پسین خواهد بود. در رابطه (۴)، $\frac{N_{ijh}}{N_{.jh}}$ ها و m_{jh} ها به ترتیب وزن ها و برآوردهای نارایب مستقیم طبقه‌های پسین هستند که به وسیله متغیر کمکی تشکیل شده‌اند.

۳-۲. برآورد مرکب^۲

برآوردگر مرکب، ترکیب خطی از برآوردگرهای ترکیبی و مستقیم ناحیه کوچک است. این برآوردگر دارای شکل کلی زیر است:

$$\hat{P}_i^{SH} = W_i \hat{P}_i^D + (1 - W_i) \hat{P}_i^S \quad (6)$$

که در آن، \hat{P}_i^D برآوردگر مستقیم (نسبت) مبتنی بر طرح برای نامین (نسبت) ناحیه کوچک است و $0 \leq W_i \leq 1$. مهمترین مسئله، تعیین وزن مناسبی است که توازن بین اریبی برآوردگر ترکیبی و بی‌ثباتی برآوردگر مستقیم را برقرار کند. وزن‌های مختلفی در متون پیشنهاد شده اند در اینجا، وزن مناسب که با مینیمم کردن میانگین توان دوم خطا (MSE) نسبت به W_i به دست می‌آید، ذکر می‌شود. میانگین توان دوم خطا برای رابطه (۶) با فرض $Cov(\hat{P}_i^D, \hat{P}_i^S) \cong 0$ به صورت رابطه (۷) است:

1 • Synthetic Estimator

2 • Composite Estimator

۳. این فرض می‌تواند برقرار باشد، زیرا، تجربه‌های عملی نشان داده‌است که تغییرات \hat{P}_i^S نسبت به \hat{P}_i^D بسیار کمتر بوده به طوری که گویا \hat{P}_i^S در مقایسه با \hat{P}_i^D ثابت است.

$$\begin{aligned}
 \text{MSE}(\hat{P}_i^{SH}) &= E(\hat{P}_i^{SH} - P_i)^2 & (7) \\
 &= E\{W_i(\hat{P}_i^D - P_i) + (1 - W_i)(\hat{P}_i^S - P_i)\}^2 \\
 &= W_i^2 \text{Var}(\hat{P}_i^D) + (1 - W_i)^2 \text{MSE}(\hat{P}_i^S)
 \end{aligned}$$

بنابراین، وزن بهینه به صورت زیر به دست می‌آید:

$$W_i^{opt} = \frac{\text{MSE}(\hat{P}_i^S)}{\text{MSE}(\hat{P}_i^S) + \text{Var}(\hat{P}_i^D)} \quad (8)$$

باید توجه کرد که وقتی برآوردگر مستقیم قابل اعتماد باشد، وزن بیشتری خواهد داشت، در غیر این صورت وزن بیشتری به برآوردگر ترکیبی داده خواهد شد. در عمل، وزن بهینه در رابطه (۸) نیز، اغلب پس از برآورد $\text{MSE}(\hat{P}_i^S)$ و $\text{Var}(\hat{P}_i^D)$ برآورد می‌شود. بر طبق طرح نمونه‌گیری چنانچه، داده‌های نمونه از روی n امین ناحیه کوچک موجود باشد، یک برآوردگر نارایب مبتنی بر مدل برای $\text{Var}(\hat{P}_i^D)$ نیز موجود خواهد بود. بنابراین، کافی است برآوردگری را برای $\text{MSE}(\hat{P}_i^S)$ به دست آورد:

$$\begin{aligned}
 \text{SE}(\hat{P}_i^S) &= E(\hat{P}_i^S - P_i)^2 & (9) \\
 &= E(\hat{P}_i^S - \hat{P}_i^D + \hat{P}_i^D - P_i)^2 \\
 &= E(\hat{P}_i^S - \hat{P}_i^D)^2 + E(\hat{P}_i^D - P_i)^2 + 2E[(\hat{P}_i^S - \hat{P}_i^D)(\hat{P}_i^D - P_i)]
 \end{aligned}$$

چنانچه $\text{Cov}(\hat{P}_i^D, \hat{P}_i^S) \cong 0$ خواهیم داشت:

$$\begin{aligned}
 E[(\hat{P}_i^S - \hat{P}_i^D)(\hat{P}_i^D - P_i)] &= E[(\hat{P}_i^S - P_i + P_i - \hat{P}_i^D)(\hat{P}_i^D - P_i)] & (10) \\
 &= -\text{Var}(\hat{P}_i^D)
 \end{aligned}$$

در نتیجه، رابطه (۱۱) به دست می‌آید:

$$\text{MSE}(\hat{P}_i^S) = E(\hat{P}_i^S - \hat{P}_i^D)^2 - \text{Var}(\hat{P}_i^D) \quad (11)$$

بدین ترتیب، می‌توان رابطه (۱۱) را براساس رابطه زیر که ممکن است منفی باشد به طور نااریب برآورد کرد:

$$M\hat{S}E(\hat{P}_i^S) = (\hat{P}_i^S - \hat{P}_i^D)^2 - \hat{V}ar(\hat{P}_i^D) \quad (12)$$

با قرار دادن رابطه (۱۲) و $\hat{V}ar(\hat{P}_i^D)$ به ترتیب به جای $MSE(\hat{P}_i^S)$ و $Var(\hat{P}_i^D)$ در رابطه (۸)، برآوردی از وزن بهینه که وابسته به داده‌ها است، به دست می‌آید. با این وجود، این وزن می‌تواند از قبل تعیین شده و یا وابسته به اندازه نمونه باشد. خواننده برای مشاهده سایر انواع وزن‌ها به کوپاس^۱ (۱۹۷۲)، گوش و رانو^۲ (۱۹۹۴)، و تامپسون و هولموی (۱۹۹۸) ارجاع داده می‌شود. پرسل و کیش (۱۹۷۹) نیز، وزن مشترکی را برای کلیه ناحیه‌های کوچک پیشنهاد کرده‌اند. چنانچه نواحی کوچک خیلی متفاوت باشند، این وزن مشترک ممکن است به طور مطلوب عمل نکند. شیل (۱۹۷۸) شیوه برآورد وزن در رابطه (۶) را از طریق مینیم کردن MSE بررسی کرده است. وی، شرایطی را که تحت آن برآوردگر مرکب منجر به کاهش قابل ملاحظه‌ای در MSE می‌شود، ارائه داده است. تابع زیر را در نظر بگیرید:

$$f(W_i) = \frac{MSE(\hat{P}_i^{SH})}{MSE(\hat{P}_i^S)} = (R_i + 1)W_i^2 - 2W_i + 1 \quad (13)$$

که در آن $R_i = \frac{Var(\hat{P}_i^D)}{MSE(\hat{P}_i^S)}$ ، این تابع دارای ویژگی مطلوب زیر است:

$$SE(\hat{P}_i^{SH}) < MSE(\hat{P}_i^S) \Leftrightarrow f(W_i) < 1 \Leftrightarrow (R_i + 1)W_i - 2 < 0 \Leftrightarrow W_i < \frac{2}{R_i + 1} \quad (14)$$

1. Copas

2. Thompson & Holmoy

این ویژگی به بازه‌ای به صورت $\left(0, \frac{2}{R_i + 1}\right)$ برای W_i می‌رسد که در آن برآوردگر مرکب لزوماً بهتر از هر دو مؤلفه خود عمل می‌کند. برای مثال، چنانچه $R_i = 1$ باشد به ازای تمام $W_i \in (0, 1)$ ، $MSE(\hat{P}_i^{SH})$ کوچکتر از $MSE(\hat{P}_i^S)$ و $Var(\hat{P}_i^D)$ است. دو روشی که تا کنون در نظر گرفته شده‌اند تحت رهیافت فراوانی‌گرا^۱ قرار دارند. اکنون روش بیز تجربی (EB) معرفی می‌شود.

۲-۳. برآوردگر بیز تجربی^۲

ابتدا، جامعه را به ناحیه‌های کوچک مختلف $1, 2, \dots, L$ که L معلوم است تقسیم می‌کنیم. فرض می‌شود هدف، برآوردنسبت واقعی جامعه برای ناحیه کوچک i ام (P_i) است ($i=1, 2, \dots, L$). برآوردگر مستقیم (P_i) از نمونه موجود در نمونه‌گیری کشوری در دسترس است، همچنین، بردار اطلاعات کمکی $x_i = (x_{i1}, \dots, x_{is})'$ از نمونه‌گیری کشوری یا داده‌های ثبتی و یا سرشماری‌های گذشته در اختیار است. باید توجه داشت که این اطلاعات کمکی با P_i مرتبط هستند.

تبدیل g که تابعی از یک متغیر ساده و دارای مشتق مرتبه اول پیوسته و غیر صفر است، در نظر گرفته می‌شود:

$$g_i = g(P_i) \quad i = 1, 2, \dots, L$$

فرض می‌شود \hat{g}_i برآورد g_i است. اکنون، مدل ناحیه کوچک زیر در نظر گرفته می‌شود:
الف) با شرط گذاری روی $g = (g_1, g_2, \dots, g_L)'$ ، \hat{g}_i ها دوبدو مستقل هستند، و برای اعضای که به ناحیه کوچک i ام تعلق دارند، $E(\hat{g}_i | g_i) = g_i$ و $V(\hat{g}_i | g_i) = \delta_i^2 (> 0)$ که در آن δ_i^2 ها مقادیر معلوم هستند.
ب) g_i ها مستقل و هم توزیع هستند با توزیع $N(x_i' \beta, \tau^2)$.

این مدل را به صورت زیر نیز می‌توان نوشت:

-
- 1 • Frequentist
 - 2 • Empirical Bayes Estimator

$$\hat{g} = X\beta + t + e \quad (15)$$

که در آن \hat{g} ، t و e بردارهای $(L \times 1)$ ، t بردار اثرهای تصادفی ناحیه ای، e بردار خطاهای تصادفی نمونه‌گیری، و \hat{g} دارای یک توزیع نرمال چند متغیره است. ماتریس X ماتریس طرح $(L \times s)$ و β بردار $(s \times 1)$ از پارامترهای مجهول است. فرض می‌کنیم بردارهای t و e مستقل و \sum و ∇ ماتریس‌های قطری بوده که مؤلفه‌های روی قطر آنها به ترتیب τ^2 و δ_i^2 هستند. همچنین

$$E(e|\hat{g}) = 0, \quad \text{Var}(e|\hat{g}) = \nabla$$

و

$$t \sim N(0, \Sigma)$$

مؤلفه‌های واریانس δ_i^2 به وسیله فرمول‌های مناسب واریانس نمونه‌گیری کشوری به دست می‌آیند. در این بخش، تبدیل لوژستیک را به صورت زیر تعریف می‌کنیم:

$$\hat{g}_i = \text{Ln} \frac{\hat{P}_i}{1 - \hat{P}_i} \quad i = 1, 2, \dots, L$$

برآورد بیز \hat{g}_i به شکل زیر است:

$$\hat{g}_i^B = (1 - B_i)\hat{g}_i + B_i x_i' \beta; \quad i = 1, 2, \dots, L \quad (16)$$

که در آن x_i' ، i امین سطر ماتریس X بوده و:

$$B_i = \frac{\delta_i^2}{\delta_i^2 + \tau^2} \quad (17)$$

ضریب $(1 - B_i)$ میزان عدم قطعیت در الگوسازی θ_i (یعنی نسبت واریانس (δ_i^2) به واریانس کل $(\delta_i^2 + \tau^2)$) را اندازه می‌گیرد. بنابراین، برآوردگر بیز تجربی نسبت تغییرات بین ناحیه‌های کوچک را به دقت برآوردگر مستقیم به درستی مورد استفاده قرار می‌دهد. در ضمن، این برآوردگر

برای عموم طرح‌های نمونه‌گیری معتبر است، زیرا، به جای تک تک اعضای جامعه، تنها P_i ها الگوسازی شده‌اند.

در تحلیل بیز تجربی τ^2 و β نامعلوم هستند و بایستی آنها را به وسیله داده ها برآورد کرد. این برآوردها با حل تکراری معادلات زیر به دست می‌آیند:

$$\begin{cases} \tau^2 = (g - X\beta)'V^{-1}(g - X\beta)/L - s \\ \beta = (X'V^{-1}X)^{-1}X'V^{-1}g \end{cases} \quad (18)$$

که در آن، V ماتریس قطری $(s \times s)$ است که درایه (i, i) آن برابر $\delta_i^2 + \tau^2$ است. پس از اینکه τ^2 با حل تکراری معادلات بالا به دست آمد، بهترین برآوردگر نارایب خطی (BLUE)، از رابطه زیر به دست می‌آید:

$$\hat{\beta} = (X'U^{-1}X)X'U^{-1}\hat{g}$$

که در آن، U ماتریس قطری $(s \times s)$ است که از ماتریس V با قرار دادن $\hat{\tau}^2$ به جای τ^2 نتیجه می‌شود.

اکنون، با توجه به رابطه (۱۶)، برآوردگر بیز تجربی \hat{g}_i به صورت زیر است:

$$\hat{g}_i^{EB} = (1 - \hat{B}_i)\hat{g}_i + \hat{B}_i x_i' \hat{\beta} \quad i = 1, 2, \dots, L \quad (19)$$

$$\hat{B}_i = \delta_i^2 / (\delta_i^2 + \hat{\tau}^2)$$

که در آن، $\hat{B}_i = \delta_i^2 / (\delta_i^2 + \hat{\tau}^2)$. در نهایت، برآوردگر بیز تجربی \hat{P}_i^{EB} ، برای مدل لوژستیک از رابطه زیر به دست می‌آید:

$$\hat{P}_i^{EB} = \frac{e^{\hat{g}_i^{EB}}}{1 + e^{\hat{g}_i^{EB}}} \quad i = 1, 2, \dots, L \quad (20)$$

در اینجا به این بحث خاتمه می‌دهیم و خواننده برای آگاهی از جزئیات بیشتر به چاند و آلکساندر^۱ (۱۹۹۵) ارجاع داده می‌شود.

۳. یک کاربرد

در ایران، ۲۸ استان وجود دارد. در سال‌های اخیر به منظور برآورد نسبت‌های بیکاری در سطح استان‌ها، مرکز آمار ایران طرح‌های نمونه‌گیری مجزا برای هر استان به صورت سالانه طراحی کرده است. برای اجرای این طرح‌های نمونه‌گیری، در کل کشور بیش از ۱۰۰۰۰۰ نمونه انتخاب می‌شود. هزینه ارایه برآورد نسبت بیکاری با این روش به مرکز آمار ایران اجازه نمی‌دهد که این برآوردها را ماهانه یا حداقل به صورت فصلی ارایه کند. از آنجا که این اطلاعات به صورت ماهانه و یا فصلی مورد نیاز اقتصاددانان و پژوهشگران است و با توجه به محدودیت‌های اجرایی، بودجه‌ای، و... و طولانی شدن زمان تهیه و اطلاع‌رسانی آن، به نظر می‌رسد استفاده از روش‌های برآورد ناحیه کوچک امکان مناسبی را برای تأمین این نیاز مهیا کند، لذا، برای برآورد این نسبت در سال ۱۳۸۰ از روش‌های برآورد ناحیه کوچک، معرفی شده در بخش قبل، استفاده شد تا با مقایسه عملکرد آنها به پیشنهادی در کارگیری یک و یا چند روش دست یابیم. این کاربرد در زیر توضیح داده می‌شود.

طرح نمونه‌گیری برای برآورد نسبت بیکاری در سطح ملی برای سال ۱۳۸۰ طراحی شد. بر این مبنای نمونه‌ای در حدود ۱۳۰۰۰ در سطح کل کشور انتخاب شد. نظر به اینکه نمونه‌ها بین تمام ۲۸ استان براساس اندازه آنها تخصیص داده شدند، لذا تعدادی نمونه از هریک از استان‌ها وجود داشت. این امر، ما را قادر ساخت تا برآورد مستقیمی از نسبت بیکاری در هر استان داشته باشیم. افزون بر این، طرح‌های نمونه‌گیری مجزا، همانند سال‌های پیش نیز برای هر استان به منظور ارایه برآوردهای مبتنی بر طرح نسبت‌های بیکاری طراحی شد. بنابراین، دو برآورد یعنی برآورد مستقیم و برآورد مبتنی بر طرح را برای هر استان قبل از بهره‌گیری از روش‌های برآورد ناحیه کوچک، در اختیار داشتیم. برآوردهای مبتنی بر طرح در ستون دوم جدول (۱) نشان داده شده‌اند. این برآوردها که با بهینه کردن نمونه‌ها در سطح استان‌ها برای اردیبهشت ۱۳۸۰ (تعداد نمونه ۱۰۴۵۶۶) محاسبه شده‌اند،^۲ مبنای مقایسه برآوردهای ناحیه کوچک هستند.

1. Chand & Alexander

۲. منبع: نشریه آمارگیری از ویژگی‌های اشتغال و بیکاری خانوار- اردیبهشت ۱۳۸۱ - مرکز آمار ایران

برآورد ترکیبی که در ستون سوم جدول (۱) به نمایش در آمده است اولین برآورد غیرمستقیمی است که با استفاده از روش‌های برآورد ناحیه کوچک به دست آمد. وسیله به کار رفته برای ارایه این برآوردها بر حسب چهار گروه سنی صورت گرفته است زیرا انتظار می رود که متغیر سن تا حدودی با وضعیت بیکاری مرتبط باشد. وزن $\frac{N_{ijh}}{N_{.jh}}$ در فرمول (۴) از روی داده‌های سرشماری سال ۱۳۷۵ محاسبه شد. همان طور که ملاحظه می شود خطای مطلق (قدرمطلق تفاوت دو برآورد) بین برآورد ترکیبی و برآورد مبتنی بر طرح از ۰/۰۰۰۹ برای استان مازندران تا ۰/۱۶۶۹۹ برای استان سیستان و بلوچستان تغییر می‌کند. شرایطی که تحت آن برآوردهای ترکیبی رضایت بخش خواهند بود به شرح زیر است:

۱. مناسب بودن متغیری که طبقه‌ها (طبقه‌های پسین) را تشکیل می‌دهد. چنانچه این متغیر در طبقات تغییرات کمی ایجاد کند، به نظری رسد که برآوردگر ترکیبی مناسب‌تر باشد.
۲. استفاده از وزن‌های بهنگام. چنانچه داده‌های ثبتي مرتبطی وجود داشته باشد، می‌توان برای برآوردگر ترکیبی وزن‌های بهنگام‌تری را به کار گرفت.

این دو مشکل، استفاده از برآوردگر ترکیبی را محدود می‌کند. بنابراین، به جای تمرکز روی این برآوردگر، آنرا برای ارایه برآورد مرکب که در ستون چهارم جدول (۱) نشان داده شده است، به دست آوردیم. در این حالت، خطای مطلق از ۰/۰۰۰۸۲ برای استان ایلام تا ۰/۰۴۲۲ برای استان هرمزگان تغییر می‌کند.

برای محاسبه برآورد بیز تجربی (EB) نسبت بیکاری استان‌های کشور در ابتدا، ماتریس طرح X با $S = 2$ به صورت زیر تعریف می‌شود:

$$\begin{cases} x_{i1} = 1 \\ x_{i2} = N_i \end{cases}$$

که در آن، N_i جمعیت فعال اقتصادی مربوط به سال (۱۳۷۵) هر یک از استان‌های کشور است.

برای به دست آوردن برآورد δ_i^2 از روش دلنا استفاده شد. این روش، به صورت زیر است:

فرض می‌کنیم $\mathcal{Y}_1, \mathcal{Y}_2, \dots, \mathcal{Y}_n$ یک نمونه تصادفی و $\hat{\theta}_n$ برآورد پارامتر مورد نظر باشد به طوری که برای n بزرگ داشته باشیم:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma^2)$$

آن گاه، برای تابعی مانند $g(\cdot)$ با مقدار حقیقی و مشتق مرتبه اول پیوسته در همسایگی θ داریم:

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \xrightarrow{d} N(0, g'^2(\theta)\sigma^2)$$

که در آن، $\sigma^2 = \text{var}(\hat{\theta}_n)$.

برآورد واریانس (σ^2) برای نسبت بیکاری با توجه به اینکه نمونه‌ها از طریق نمونه‌گیری تصادفی ساده به دست آمده‌اند، از رابطه زیر به دست می‌آید:

$$\hat{\sigma}^2 = \text{var}(p_i^D) = \frac{P_i^D(1 - P_i^D)}{n_i - 1}$$

که در آن n_i تعداد نمونه تخصیص یافته به استان نام و مقدار P_i^D برآورد مستقیم نسبت بیکاری در استان نام است.

با توجه به برآورد $\hat{\sigma}^2$ ، مقدار $\hat{\delta}_i^2$ به صورت زیر برآورد می‌شود:

$$\hat{\delta}_i^2 = g'^2(\hat{P}_i)\hat{\sigma}^2$$

برای برآورد $\hat{\tau}^2$ از روش حل تکراری معادله (۱۸) و برنامه نوشته شده در محیط برنامه‌نویسی SAS/IML بهره‌گرفته شده است. برآورد $\hat{\tau}^2$ برای مسئله برآورد بیز تجربی نسبت بیکاری استان‌های کشور در سال ۱۳۸۰ برابر است با:

$$\hat{\tau}^2 = ۰.۰۵۵۹۶$$

و برآورد بردار پارامترها به صورت زیر است:

$$\hat{\beta} = \begin{pmatrix} -2.07 \\ -1.2E-7 \end{pmatrix}$$

ستون آخر جدول (۱) به برآوردهای بیز تجربی (EB) اختصاص داده شده است. این برآوردها پس از تقریباً ۵۰۰۰ بار حل تکراری معادله‌های (۱۸) با استفاده از برنامه SAS/IML محاسبه شده‌اند. متغیر کمکی که در مدل (۱۵) به کار رفته است جمعیت فعال اقتصادی^۱ است. همان گونه که ملاحظه می‌شود خطای مطلق از ۰/۰۰۰۶۸ برای استان گیلان تا ۰/۰۴۰۵۳ برای استان هرمزگان تغییر می‌کند.

برای ارزیابی سه برآوردگر یادشده، میانگین قدر مطلق خطا^۱ که در سطر آخر جدول (۱) مشخص شده، محاسبه شده‌است. این معیار، نشان می‌دهد که برآوردهای مرکب و EB بهتر از برآوردهای ترکیبی عمل می‌کنند زیرا، در مقایسه با برآوردهای ترکیبی، آنها به برآوردهای مبتنی بر طرح (برآورد استانی) نزدیکتر هستند. برآوردهای مرکب و EB خیلی به هم نزدیک هستند. البته، این بررسی دقیقاً نشان نمی‌دهد که برآوردگر EB بهبود مناسبتری نسبت به برآوردگر مرکب به دست می‌دهد. براساس یافته‌های این پژوهش، هم برآورد بیز تجربی و هم برآورد مرکب را می‌توانیم به طور یکسان برای برآورد نسبت بیکاری در سطح استان‌ها بدون اینکه یکی بر دیگری برتری داشته باشد، به کار ببریم.

$$2. ARE = 100 \left[\frac{1}{L} \sum_{i=1}^L |p_i - P_i^{Des}| / |P_i^{Des}| \right] \quad \text{که } P_i^{Des} \text{ برآورد مبتنی بر طرح است}$$

جدول-۱. برآوردهای نسبتهای بیکاری در سال ۱۳۸۰ در استانهای^۱ کشور

استان	طرح	ترکیبی	مرکب	بیزنجری
اردبیل	۰/۰۹۹۹۶	۰/۱۴۷۸۹	۰/۱۱۷۴۶	
بوشهر	۰/۱۷۰۷۵	۰/۱۰۴۹۱		۰/۱۷۸۸۰
چهارمحال و بختیاری	۰/۰۹۱۱۱		۰/۰۶۳۵۱	۰/۰۶۶۵۹
آذربایجان شرقی	۰/۰۸۷۵۲		۰/۰۸۳۱۷	۰/۰۷۸۵۷
اصفهان	۰/۱۳۸۷۰	۰/۱۱۵۳۳		۰/۱۵۲۶۷
فارس	۰/۱۵۸۰۰	۰/۱۴۴۲۱	۰/۱۴۰۰۳	
قزوین	۰/۱۳۹۸۷	۰/۰۹۱۷۸	۰/۱۱۱۱۹	
قم	۰/۱۰۵۳۲	۰/۰۷۴۱۶		۰/۰۷۳۰۸
گیلان	۰/۱۳۴۹۳	۰/۲۰۵۷۵	۰/۱۳۶۰۹	
گلستان	۰/۱۴۸۱۸	۰/۱۷۰۱۲		۰/۱۳۱۴۳
همدان	۰/۱۰۷۷۴	۰/۱۲۱۶۵		۰/۰۹۶۰۰
هرمزگان	۰/۲۲۲۹۶	۰/۱۱۱۱۱۳	۰/۲۶۵۱۶	
ایلام	۰/۱۵۰۲۶	۰/۲۴۴۹۵		۰/۱۴۲۳۲
کرمان	۰/۱۵۲۶۵	۰/۱۲۰۹۸	۰/۱۳۰۸۲	
کرمانشاه	۰/۱۷۵۶۳	۰/۲۶۴۹۸		۰/۱۵۱۰۸
خراسان	۰/۱۳۶۳۷	۰/۰۹۰۸۴	۰/۱۲۹۶۶	
خوزستان	۰/۱۹۸۰۵	۰/۲۵۷۳۷	۰/۲۲۷۲۸	
کهگیلویه و بویراحمد	۰/۲۰۰۳۸	۰/۲۱۶۸۱	۰/۲۲۸۰۷	
کردستان	۰/۱۲۱۶۹		۰/۱۳۰۶۵	۰/۱۱۳۳۲
لرستان	۰/۳۰۴۹۰	۰/۲۶۶۲۹	۰/۲۶۶۷۳	
مرکزی	۰/۱۰۵۲۹		۰/۰۹۷۳۹	۰/۱۲۲۰۵
مازندران	۰/۱۳۱۱۴		۰/۱۲۰۲۹	۰/۱۳۷۶۴
سمنان	۰/۱۵۷۰۶	۰/۰۶۸۱۳	۰/۱۳۳۱۵	
سیستان و بلوچستان	۰/۲۸۶۸۳	۰/۱۱۹۸۴		۰/۲۷۳۷۲
تهران	۰/۱۲۱۸۷	۰/۰۸۴۴۴	۰/۱۱۹۵۲	
آذربایجان غربی	۰/۰۸۷۹۶	۰/۱۲۹۶۶		۰/۰۸۴۷۹
یزد	۰/۱۲۱۲۳	۰/۰۶۷۷۴	۰/۱۰۹۵۶	
زنجان	۰/۱۰۱۹۸		۰/۱۲۰۰۹	۰/۰۸۷۸۲
حدافل خطای مطلق ^۲	-	۰/۰۰۰۱۹۰۰	۰/۰۰۰۸۱۵۴	۰/۰۰۰۶۸۲۳
حداکثر خطای مطلق ^۳	-	۰/۱۶۶۹۹۰۰	۰/۰۴۲۱۹۸۰	۰/۰۴۰۵۳۳۴
میانگین قدر مطلق خطا ^۴	-	۰/۰۴۴۶۵	۰/۰۴۴۸۲	۰/۰۱۳۴۷

^۱ نزدیکترین برآورد به برآورد مبتنی بر طرح برجسته شده است.

^۲ این مقادیر قبل از گرد کردن برآوردها محاسبه شده اند.

منابع

- Chand, N., and Alexander, C.H. (1995). Indirect Estimation of Rates and Proportions for Small Areas with Continuous Measurement. *In Proceeding of the Section on Survey Research Methods, American Statistical Association*, 549-554.
- Copas, J. B. (1972). Empirical Bayes methods and the repeated use of a standard, *Biometrika*, 59, 349-360.
- Ghosh, M., and Rao, J.N.K. (1994). Small Area Estimation: an appraisal (with discussion). *Statistical Science*, 9, 65 - 93.
- Marker, D. A. (1999). Organization of Small Area Estimators Using a Generalized Linear Regression Framework. *Journal of Official Statistics*, 15, 1 - 24.
- Pfeffermann, D. (1999). Small Area Estimation-Big Developments. *Paper presented at the international association of survey statisticians satellite conference*, Riga, Latvia, 20-21 August 1999.
- Purcell, N. J., and Kish, L. (1979). Estimation for Small Domains. *Biometrics*, 35, 365 - 384.
- Purcell, N. J., and Kish, L. (1980). Postcensal Estimates for Local Areas (or Domains). *International Statistical Review*, 48, 3 - 18.
- SAS Institute Inc. (1990). *SAS/IML Software: Reference, Version 6*, First Edition, Cary, Nc: SAS Institute Inc.
- Schaible, W. L. (1978). Choosing Weight for Composite Estimators for Small Area Statistics. *In Proceedings of the Section on Survey Research Methods, American Statistical Association*, 741-746.
- Schaible, W.L., ed. (1995), *Lecture Notes in Statistics: Indirect Estimators in U.S. Federal Programs*, New York: Springer.

Thompson, I., and Holmoy A.M.K. (1998). Combining data from surveys and administrative record system: The Norwegian experience. *International Statistical Review*, 66, 201-221.

U.S. National Center for Health Statistics. (1968). *Synthetic State Estimations of Disability*. Washington, D.C., U.S. Public Health Service.